Workshop on epidemiological modelling: a stochastic compartmental models perspective

Lorenzo Rimella

October 24, 2023

Outline

- Introduction
  - Compartmental models for epidemiology
  - Deterministic vs Stochastic
  - Continuous in time vs discrete in time
- Practical implementation of stochastic compartmental models
- Compartmental models as hidden Markov models (HMM)
  - Likelihood computation in HMM and issues
  - Likelihood approximation through Sequential Monte Carlo (SMC)
  - Parameters inference
- Practical implementation of SMC and black-box optimizers, and ABC
- Parametric approximation in compartmental models
  - Kalman-Filter
  - Beyond the Gaussian approximation (e.g. MAL and PAL)
  - Dealing with overdispersion
- Practical implementation of the PAL and gradient ascent
- Individual-based model
  - Definition, motivation and challenges
  - Likelihood computation and inference

What? Why? How?

- What?
  We want to understand the spread of a pathogen in a population.

- Why?
  We care about people and we want them to be fine.

- How?
  We use mathematical models and big computers to fit them.

What? Why? How?

- What?
  We want to understand the spread of a pathogen in a population.

- Why?
  We care about people and we want them to be fine.

- How?
  We use mathematical models and big computers to fit them.

What? Why? How?

- What?
  We want to understand the spread of a pathogen in a population.

- Why?
  We care about people and we want them to be fine.

- How?
  We use mathematical models and big computers to fit them.

What? Why? How?

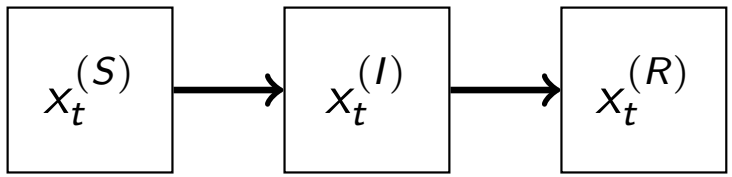
- What?
  We want to understand the spread of a pathogen in a population.

- Why?
  We care about people and we want them to be fine.

- How?
  We use mathematical models and big computers to fit them.

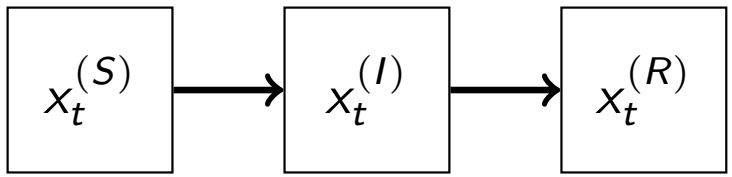
Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.



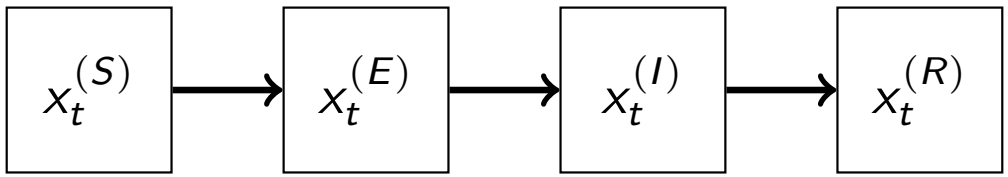
Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.
- Note that we can add as many compartments as we want!



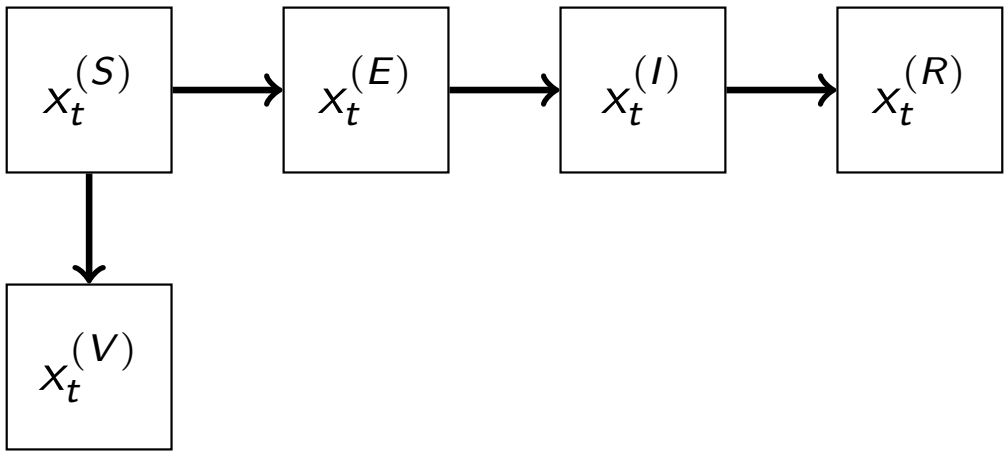
Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.
- Note that we can add as many compartments as we want!



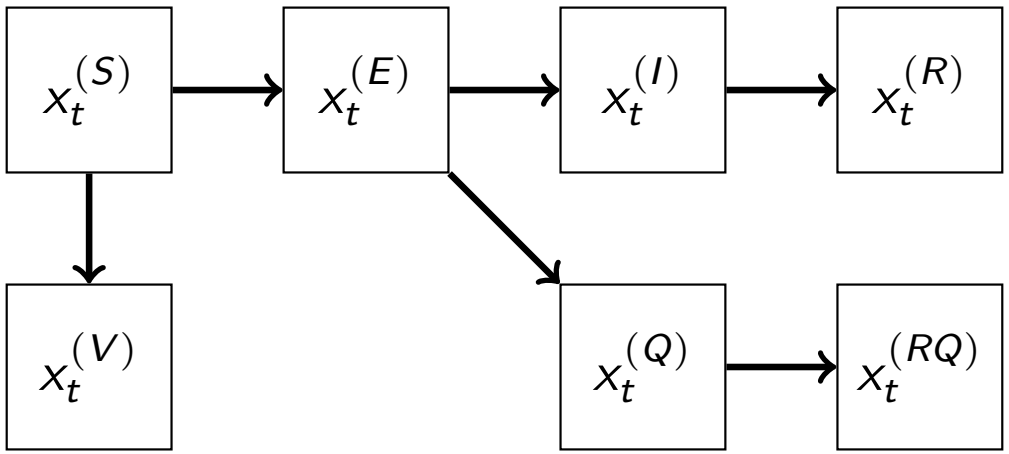
Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.
- Note that we can add as many compartments as we want!



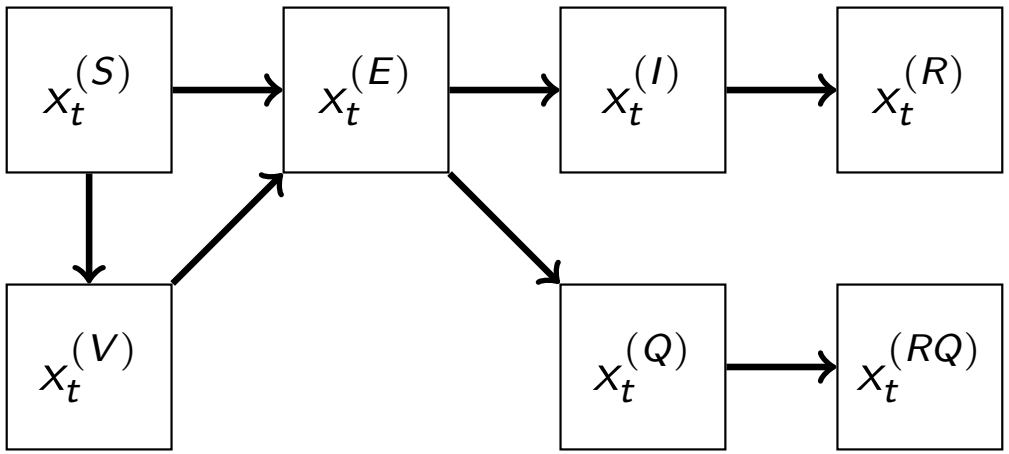
Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.
- Note that we can add as many compartments as we want!



Compartmental models

- We can model the spread of a pathogen in a population of N individuals by moving individuals across compartments.
- A well-known example is the SIR model and it will be the guiding example of the workshop.
- Note that we can add as many compartments as we want!



Deterministic compartmental models

- The movement across compartments is a system of ODE [Keeling and Rohani, 2011]:

$$\frac{dx_t^{(S)}}{dt} = -\beta \frac{x_t^{(S)} x_t^{(I)}}{N}$$
$$\frac{dx_t^{(I)}}{dt} = \beta \frac{x_t^{(S)} x_t^{(I)}}{N} - \gamma x_t^{(I)}$$
$$\frac{dx_t^{(R)}}{dt} = \gamma x_t^{(I)}$$

where β is the transmission parameter, γ is the recovery parameter and N is the population size.

Deterministic compartmental models

- The movement across compartments is a system of ODE [Keeling and Rohani, 2011]:

$$\frac{dx_t^{(S)}}{dt} = -\beta \frac{x_t^{(S)} x_t^{(I)}}{N}$$
$$\frac{dx_t^{(I)}}{dt} = \beta \frac{x_t^{(S)} x_t^{(I)}}{N} - \gamma x_t^{(I)}$$
$$\frac{dx_t^{(R)}}{dt} = \gamma x_t^{(I)}$$

  where β is the transmission parameter, γ is the recovery parameter and N is the population size.

- We have multiple interpretations of $\beta \frac{x_t^{(S)} x_t^{(I)}}{N}$:
  - every pair of individuals meet at rate β/N;
  - susceptibles meets infectives at rate β and go infectious wp $x_t^{(I)}/N$;
  - infectives meets susceptibles at rate β and go infectious wp $x_t^{(S)}/N$.

Deterministic compartmental models

- The movement across compartments is a system of ODE [Keeling and Rohani, 2011]:

$$\frac{dx_t^{(S)}}{dt} = -\beta \frac{x_t^{(S)} x_t^{(I)}}{N}$$
$$\frac{dx_t^{(I)}}{dt} = \beta \frac{x_t^{(S)} x_t^{(I)}}{N} - \gamma x_t^{(I)}$$
$$\frac{dx_t^{(R)}}{dt} = \gamma x_t^{(I)}$$

  where β is the transmission parameter, γ is the recovery parameter and N is the population size.

- There are some advantages with this approach:
  - deterministic models can describe large epidemic dynamic;
  - simple to understand if you are a differential equations fan;
  - they simply require a numerical solution of the ODE system and a "pairing" with the observations.

Deterministic compartmental models

- The movement across compartments is a system of ODE [Keeling and Rohani, 2011]:

$$\frac{dx_t^{(S)}}{dt} = -\beta \frac{x_t^{(S)} x_t^{(I)}}{N}$$
$$\frac{dx_t^{(I)}}{dt} = \beta \frac{x_t^{(S)} x_t^{(I)}}{N} - \gamma x_t^{(I)}$$
$$\frac{dx_t^{(R)}}{dt} = \gamma x_t^{(I)}$$

  where β is the transmission parameter, γ is the recovery parameter and N is the population size.

- There are several issues with this approach:
  - they are less useful for small epidemics, even in large populations;
  - they never die out so they cannot answer questions on emerging infections or local elimination;
  - they always start in the same way;
  - do we really think epidemics are deterministic?

Stochastic compartmental models

- The movement across compartments is a system of stochastics equations:

$$
\begin{aligned}
x_{t+1}^{(S)} &= x_t^{(S)} - B_t, \\
x_{t+1}^{(I)} &= x_t^{(I)} + B_t - C_t, \\
x_{t+1}^{(R)} &= x_t^{(R)} + C_t,
\end{aligned}
$$

where $B_t \sim \text{Bin}\left(x_t^{(S)}, 1 - e^{-\beta x_t^{(I)}/N}\right)$, and $C_t \sim \text{Bin}\left(x_t^{(I)}, 1 - e^{-\gamma}\right)$.

Stochastic compartmental models

- The movement across compartments is a system of stochastics equations:

$$
\begin{aligned}
x_{t+1}^{(S)} &= x_t^{(S)} - B_t, \\
x_{t+1}^{(I)} &= x_t^{(I)} + B_t - C_t, \\
x_{t+1}^{(R)} &= x_t^{(R)} + C_t,
\end{aligned}
$$

where $B_t \sim \text{Bin}\left(x_t^{(S)}, 1 - e^{-\beta x_t^{(I)}/N}\right)$, and $C_t \sim \text{Bin}\left(x_t^{(I)}, 1 - e^{-\gamma}\right)$.

- The interpretation is straightforward:
  - individuals in a certain compartment at time t move to another compartment with a certain probability;
  - for the case S to I the movement probability depends on the number of infected in the population.

Stochastic compartmental models

- The movement across compartments is a system of stochastics equations:

$$
\begin{aligned}
x_{t+1}^{(S)} &= x_t^{(S)} - B_t, \\
x_{t+1}^{(I)} &= x_t^{(I)} + B_t - C_t, \\
x_{t+1}^{(R)} &= x_t^{(R)} + C_t,
\end{aligned}
$$

where $B_t \sim \text{Bin}\left(x_t^{(S)}, 1 - e^{-\beta x_t^{(I)}/N}\right)$, and $C_t \sim \text{Bin}\left(x_t^{(I)}, 1 - e^{-\gamma}\right)$.

- In the large population limit the deterministic and the stochastic SIR are equivalent, but we can easily add elements in the stochastic model that will break the deterministic.

Stochastic compartmental models

- The movement across compartments is a system of stochastics equations:

$$
\begin{aligned}
x_{t+1}^{(S)} &= x_t^{(S)} - B_t, \\
x_{t+1}^{(I)} &= x_t^{(I)} + B_t - C_t, \\
x_{t+1}^{(R)} &= x_t^{(R)} + C_t,
\end{aligned}
$$

where $B_t \sim \text{Bin}\left(x_t^{(S)}, 1 - e^{-\beta x_t^{(I)}/N}\right)$, and $C_t \sim \text{Bin}\left(x_t^{(I)}, 1 - e^{-\gamma}\right)$.

- There are some advantages and issues with this approach:
  - but we will let you draw your conclusions after the workshop :)

Continuous in time vs discrete in time

The considered compartmental model is discrete in time, but generally we look at epidemics from a continuous time perspective.

- Here we model the events in the system, like a new infection or a new recovery.
- Each of these events is model as an **Exp** random variable.
- Even just simulating from this model depends on the population size N.
- If we consider very small time intervals in the discrete time case we go back to the continuous time.

Continuous in time vs discrete in time

The considered compartmental model is discrete in time, but generally we look at epidemics from a continuous time perspective.

- Here we model the events in the system, like a new infection or a new recovery.
- Each of these events is model as an **Exp** random variable.
- Even just simulating from this model depends on the population size N.
- If we consider very small time intervals in the discrete time case we go back to the continuous time.
- Data arrive at discrete time anyway, why bother? :)

SIR from another perspective

Let ι_t^n be the compartment of individual n at time t then x_t is given by:

$$x_t^{(i)} = \sum_{n=1}^{N} \mathbb{I}[\iota_t^n = i], \quad \text{for } i = S, I, R.$$

For each $n = 1, \ldots, N$:

$$\iota_t^n \Big| \left(\iota_{t-1}^n\right)_{n=1,\ldots,N} \sim \textbf{Cat}\left(K_{x_{t-1}}^{(\iota_{t-1}^n, \cdot)}\right)$$

where:

$$K_x = \begin{pmatrix} e^{-\beta \frac{x^{(I)}}{N}} & 1 - e^{-\beta \frac{x^{(I)}}{N}} & 0 \\ 0 & e^{-\gamma} & 1 - e^{-\gamma} \\ 0 & 0 & 1 \end{pmatrix}.$$

K_x is the individuals' kernel and depends on the current state of the population.

SIR from Multinomial kernel

- Sum of categorical random variables with the same probability parameters is a multinomial.
- We can then write our SIR as via a multinomial kernel formulation:

$$x_0 \sim \textbf{Multi}(N, p_0)$$
$$x_t^{(i)} = \sum_{j=S,I,R} Z_{t-1,t}^{(j,i)} \text{ for } i = S, I, R$$
$$\text{with } Z_{t-1,t}^{(i,\cdot)} \sim \textbf{Multi}(x_{t-1}^{(i)}, K_{x_{t-1}}^{(i,\cdot)}) \text{ for } i = S, I, R$$

SIR from Multinomial kernel

- Sum of categorical random variables with the same probability parameters is a multinomial.
- We can then write our SIR as via a multinomial kernel formulation:

$$x_0 \sim \textbf{Multi}(N, p_0)$$
$$x_t^{(i)} = \sum_{j=S,I,R} Z_{t-1,t}^{(j,i)} \text{ for } i = S, I, R$$
$$\text{with } Z_{t-1,t}^{(i,\cdot)} \sim \textbf{Multi}(x_{t-1}^{(i)}, K_{x_{t-1}}^{(i,\cdot)}) \text{ for } i = S, I, R$$

Remark:

Obviously we do not observe the number of individuals in each compartment. The final block of our epidemic model is then the link between our compartments and the observations:

$$y_t^{(i)} \sim \textbf{Bin}(x_t^{(i)}, q^{(i)}), \text{ with } i = S, I, R.$$

General compartmental model

- We can generalise to M compartments.
- The dynamics is now:

$$x_0 \sim \textbf{Multi}(N, p_0)$$

$$x_t^{(i)} = \sum_{j=1}^{M} Z_{t-1,t}^{(j,i)} \text{ for } i = 1, \ldots, M$$

$$\text{with } Z_{t-1,t}^{(i,\cdot)} \sim \textbf{Multi}(x_{t-1}^{(i)}, K_{x_{t-1}}^{(i,\cdot)}) \text{ for } i = 1, \ldots, M.$$

- The observations are now:

$$y_t^{(i)} \sim \textbf{Bin}(x_t^{(i)}, q^{(i)}), \text{ with } i = 1, \ldots, M.$$

Practical implementation of stochastic compartmental models

Hidden Markov models

A Hidden Markov model (HMM) is a doubly stochastic process $(x_t, y_t)_{t \geqslant 1}$ where [Chopin and Papaspiliopoulos, 2020]:

- x_t is a hidden Markov chain;
- y_t are observed and conditionally independent given x_t.

To define an HMM we need three quantities:

- initial distribution $p(x_0|\theta)$
- transition kernel $p(x_t|x_{t-1}, \theta)$
- emission distribution $p(y_t|x_t, \theta)$

A model that look like this...

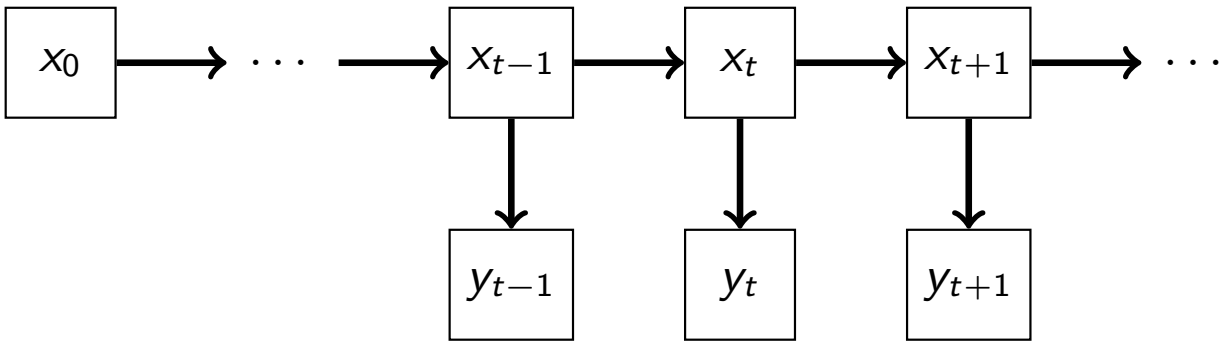
Hidden Markov models

A Hidden Markov model (HMM) is a doubly stochastic process $(x_t, y_t)_{t \geqslant 1}$ where [Chopin and Papaspiliopoulos, 2020]:

- x_t is a hidden Markov chain;
- y_t are observed and conditionally independent given x_t.

To define an HMM we need three quantities:

- initial distribution $p(x_0|\theta)$
- transition kernel $p(x_t|x_{t-1}, \theta)$
- emission distribution $p(y_t|x_t, \theta)$



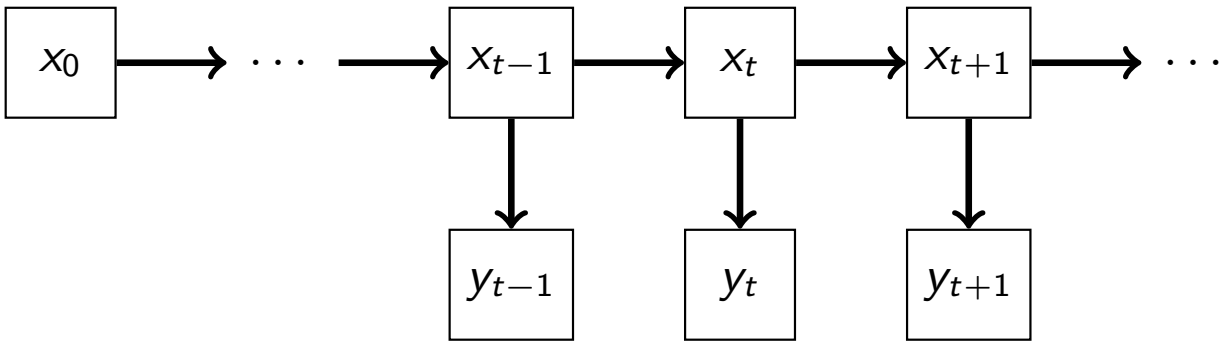
Hidden Markov models

A Hidden Markov model (HMM) is a doubly stochastic process $(x_t, y_t)_{t \geqslant 1}$ where [Chopin and Papaspiliopoulos, 2020]:

- x_t is a hidden Markov chain;
- y_t are observed and conditionally independent given x_t.

To define an HMM we need three quantities:

- initial distribution $p(x_0|\theta)$ (e.g. initial compartments);
- transition kernel $p(x_t|x_{t-1}, \theta)$
- emission distribution $p(y_t|x_t, \theta)$



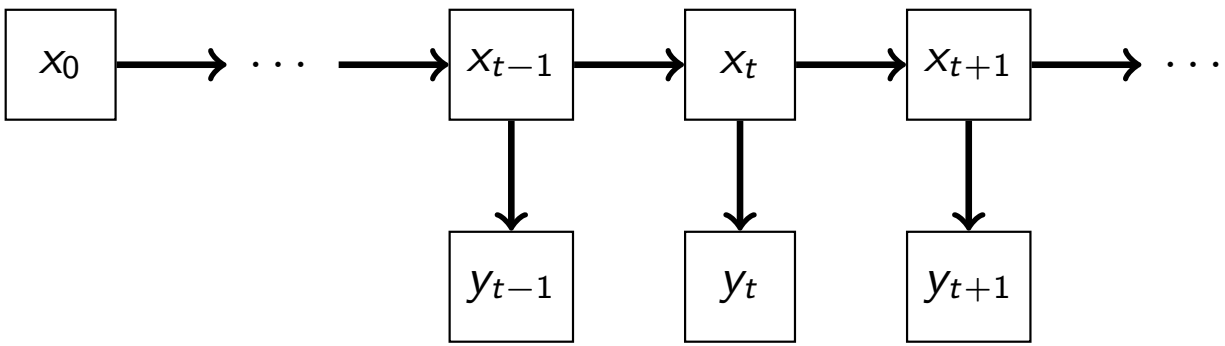
Hidden Markov models

A Hidden Markov model (HMM) is a doubly stochastic process $(x_t, y_t)_{t \geqslant 1}$ where [Chopin and Papaspiliopoulos, 2020]:

- x_t is a hidden Markov chain;
- y_t are observed and conditionally independent given x_t.

To define an HMM we need three quantities:

- initial distribution $p(x_0|\theta)$ (e.g. initial compartments);
- transition kernel $p(x_t|x_{t-1}, \theta)$ (e.g. movements across compartments);
- emission distribution $p(y_t|x_t, \theta)$



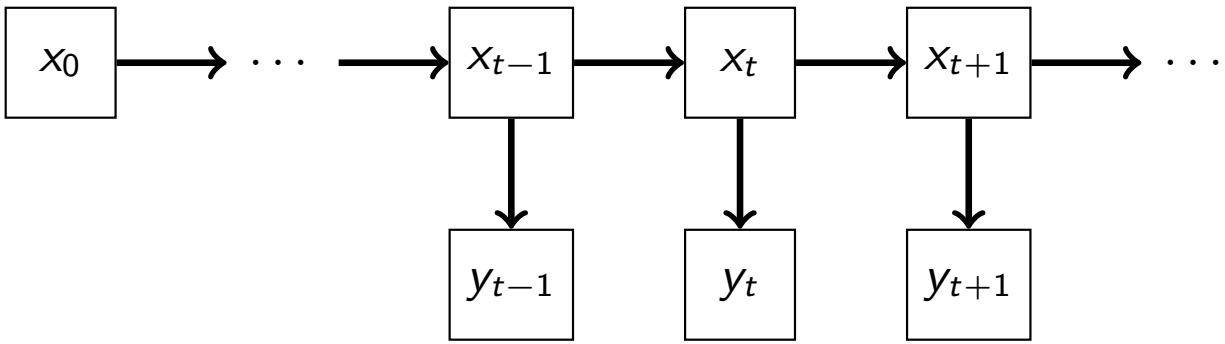
Hidden Markov models

A Hidden Markov model (HMM) is a doubly stochastic process $(x_t, y_t)_{t \geqslant 1}$ where [Chopin and Papaspiliopoulos, 2020]:

- x_t is a hidden Markov chain;
- y_t are observed and conditionally independent given x_t.

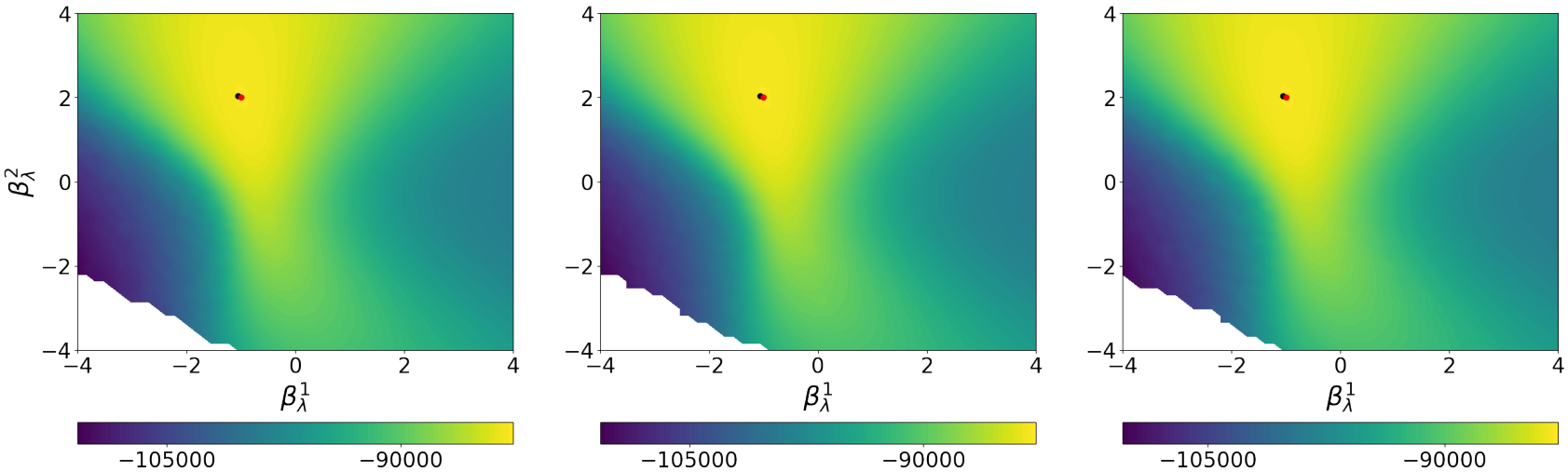
To define an HMM we need three quantities:

- initial distribution $p(x_0|\theta)$ (e.g. initial compartments);
- transition kernel $p(x_t|x_{t-1}, \theta)$ (e.g. movements across compartments);
- emission distribution $p(y_t|x_t, \theta)$ (e.g. tests on the compartments).



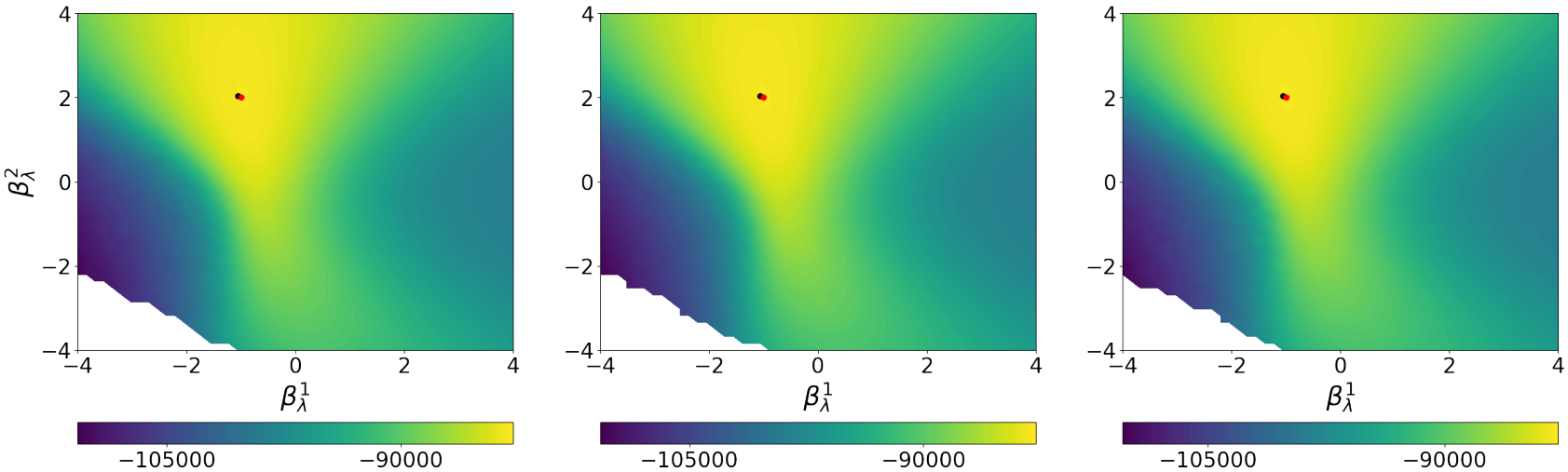
Likelihood computation

We just have a blueprint of the model, how can we find the best model given the data $y_{1:T}$?



Likelihood computation

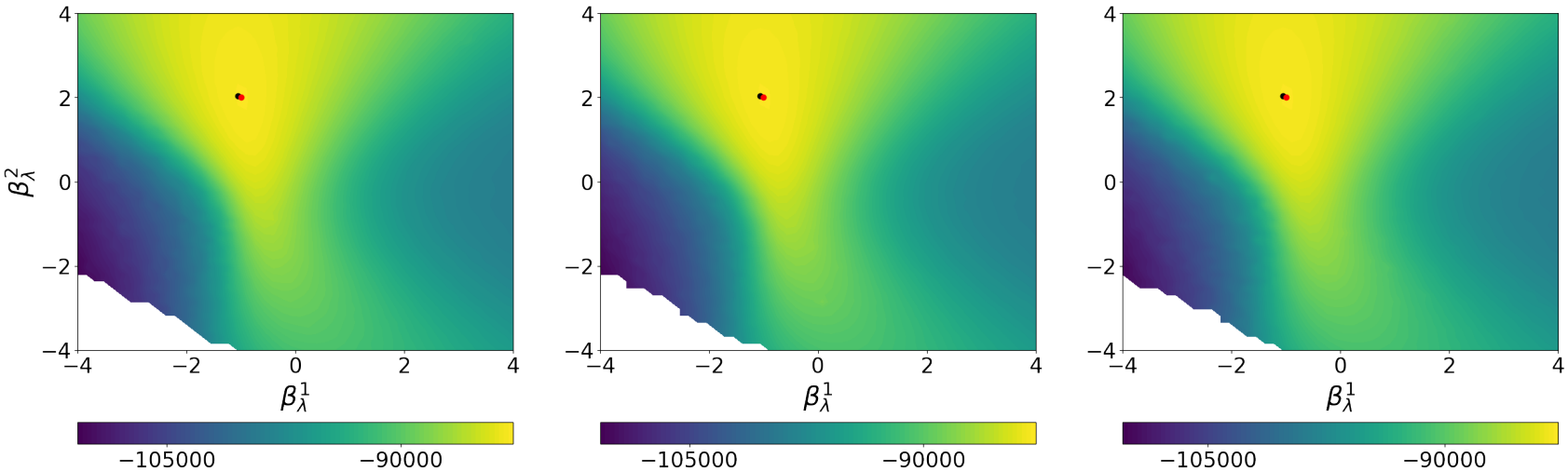
We just have a blueprint of the model, how can we find the best model given the data $y_{1:T}$? Frequentist or Bayesian



Likelihood computation

We just have a blueprint of the model, how can we find the best model given the data $y_{1:T}$?

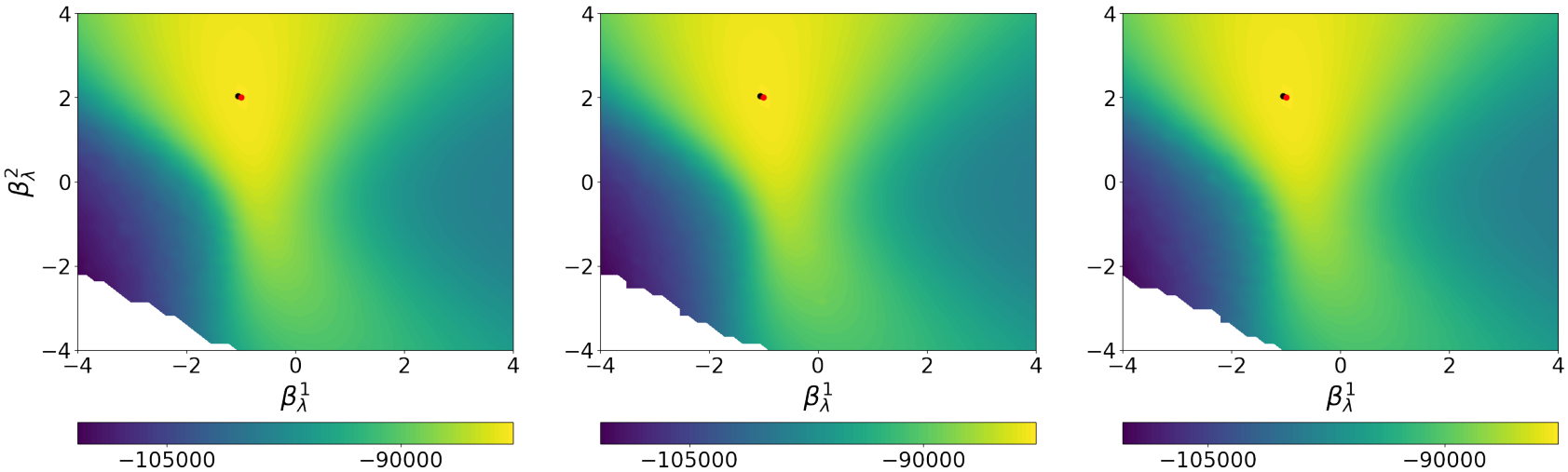
- Frequentist: find $\hat{\theta}$ that maximise $p(y_{1:T}|\theta)$.



Likelihood computation

We just have a blueprint of the model, how can we find the best model given the data $y_{1:T}$?

- Frequentist: find $\hat{\theta}$ that maximise $p(y_{1:T}|\theta)$.
- Bayesian: compute $p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)p(\theta)$.

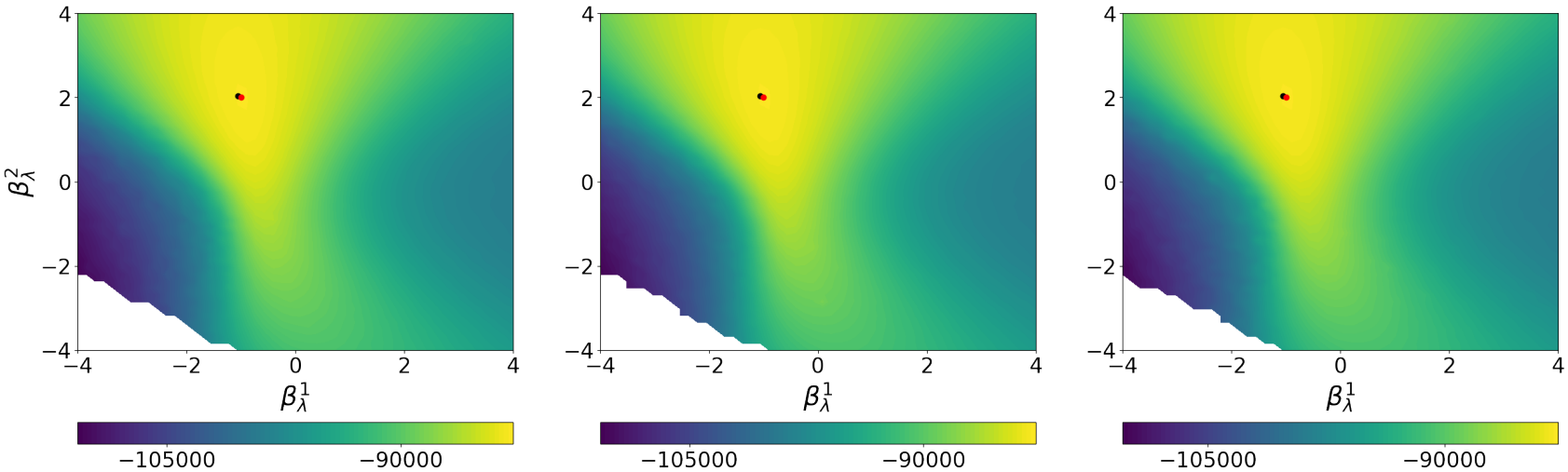


Likelihood computation

We just have a blueprint of the model, how can we find the best model given the data $y_{1:T}$?

- Frequentist: find $\hat{\theta}$ that maximise $p(y_{1:T}|\theta)$.
- Bayesian: compute $p(\theta|y_{1:T}) \propto p(y_{1:T}|\theta)p(\theta)$.

Both approaches require to be able to compute, or estimate, the likelihood $p(y_{1:T}|\theta)$.



Likelihood computation: forward algorithm

We can recursively compute the likelihood via the forward algorithm:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Likelihood computation: forward algorithm

We can recursively compute the likelihood via the forward algorithm:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Good news ✓

- The likelihood is given by the product of the likelihood increments: $p(y_{1:T}|\theta) = p(y_1|\theta)\prod_{s=2}^{t} p(y_t|y_{1:s-1},\theta)$.
- Now we can estimate perform inference as we wish.

Likelihood computation: forward algorithm

We can recursively compute the likelihood via the forward algorithm:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Good news ✓

- The likelihood is given by the product of the likelihood increments: $p(y_{1:T}|\theta) = p(y_1|\theta)\prod_{s=2}^{t} p(y_t|y_{1:s-1},\theta)$.
- Now we can estimate perform inference as we wish.

Danger ⚠

- The integrals can be computed only in some specific scenarios.

Likelihood computation for discrete HMM

In an SIR model x_t takes values on a discrete state-space.

- The integrals in the forward algorithm become sums.
- The state-space is huge:

$$\mathcal{X} = \{x \in \mathbb{N}_0 \text{ such that } x^S + x^I + x^R = N\}$$

even though some combinations have probability zero depending on the current state.

Likelihood computation for discrete HMM

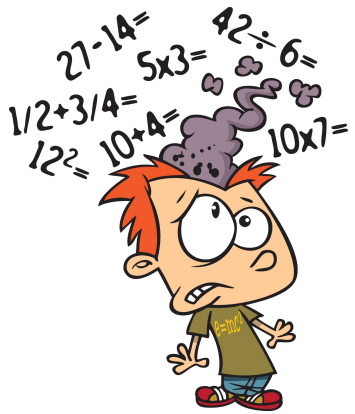
In an SIR model x_t takes values on a discrete state-space.

- The integrals in the forward algorithm become sums.
- The state-space is huge:

$$\mathcal{X} = \{x \in \mathbb{N}_0 \text{ such that } x^S + x^I + x^R = N\}$$

even though some combinations have probability zero depending on the current state.

- We do not want to perform huge sums, and even if we want we cannot when N is large.



Sequential Monte Carlo (SMC): Bootstrap Particle Filter

SMC [Chopin and Papaspiliopoulos, 2020] approximate the likelihood $p(y_{1:T}|\theta)$ via importance sampling. A well-known SMC is:

- sample $\tilde{x}_0^p \sim p(x_0|\theta)$ and set $w_0^p \leftarrow 1$;
- For $t = 1, \ldots, T$
  - resample $i_{t-1}^p \sim \textbf{Cat}\left(\frac{w_{t-1}(\theta)}{\sum_P w_{t-1}^p(\theta)}\right)$ and set $\tilde{x}_{t-1}^p = x_t^{i_{t-1}^p}$;
  - sample $x_t^p \sim p(x_t|\tilde{x}_{t-1}^p, \theta)$;
  - compute $w_t^p \leftarrow \frac{1}{P} p(y_t|x_t^p, \theta)$.

Sequential Monte Carlo (SMC): Bootstrap Particle Filter

SMC [Chopin and Papaspiliopoulos, 2020] approximate the likelihood $p(y_{1:T}|\theta)$ via importance sampling. A well-known SMC is:

- sample $\tilde{x}_0^p \sim p(x_0|\theta)$ and set $w_0^p \leftarrow 1$;
- For $t = 1, \ldots, T$
  - resample $i_{t-1}^p \sim \textbf{Cat}\left(\frac{w_{t-1}(\theta)}{\sum_P w_{t-1}^p(\theta)}\right)$ and set $\tilde{x}_{t-1}^p = x_t^{i_{t-1}^p}$;
  - sample $x_t^p \sim p(x_t|\tilde{x}_{t-1}^p, \theta)$;
  - compute $w_t^p \leftarrow \frac{1}{P} p(y_t|x_t^p, \theta)$.

The likelihood is then approximated by:

$$p(y_{1:T}|\theta) \approx \prod_{t=1}^{T} \sum_p w_t^p.$$

SMC in general

We can boost the performance via proposal distribution $\mathbf{prop}(x_t|x_{t-1})$ and resampling scheme $\mathbf{res}(i_t)$ [Whiteley and Lee, 2014, Rimella et al., 2023a]:

- sample $x_0^p \sim \mathbf{prop}(x_0)$ and set $w_0^p = \frac{p(x_0^p|\theta)}{\mathbf{prop}(x_0^p|y_{1:T})}$;
- For $t = 1, \ldots, T$
  - resample $i_{t-1}^p \sim \mathbf{res}(i_{t-1})$ and set $\tilde{x}_{t-1}^p = x_{t-1}^{i_{t-1}^p}$, $w_{t-1}^p \leftarrow \frac{w_{t-1}^{i_{t-1}^p}}{\mathbf{res}(i_{t-1}^p)}$;
  - compute $\tilde{w}_{t-1}^p = \frac{w_{t-1}^p(\theta)}{\sum_{\tilde{p}} w_{t-1}^p}$;
  - sample $x_t^p \sim \mathbf{prop}(x_t|\tilde{x}_{t-1}^p)$;
  - compute $w_t^p \leftarrow \tilde{w}_{t-1}^p \frac{p(x_t^p|\tilde{x}_{t-1}^p,\theta)p(y_t|x_t^p,\theta)}{\mathbf{prop}(x_t^p|\tilde{x}_{t-1}^p)}$.

SMC in general

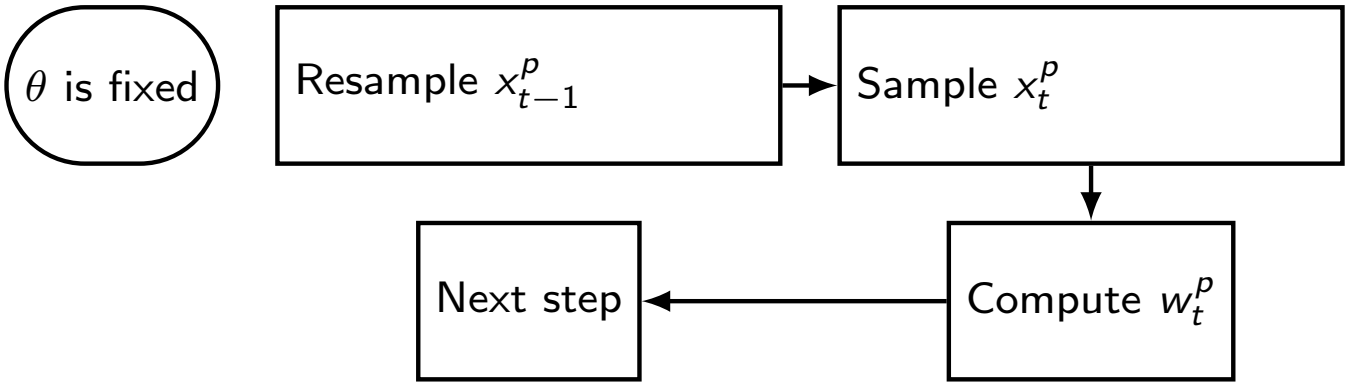
We can boost the performance via proposal distribution $\mathbf{prop}(x_t|x_{t-1})$ and resampling scheme $\mathbf{res}(i_t)$ [Whiteley and Lee, 2014, Rimella et al., 2023a]:

- sample $x_0^p \sim \mathbf{prop}(x_0)$ and set $w_0^p = \frac{p(x_0^p|\theta)}{\mathbf{prop}(x_0^p|y_{1:T})}$;
- For $t = 1, \ldots, T$
  - resample $i_{t-1}^p \sim \mathbf{res}(i_{t-1})$ and set $\tilde{x}_{t-1}^p = x_{t-1}^{i_{t-1}^p}$, $w_{t-1}^p \leftarrow \frac{w_{t-1}^{i_{t-1}^p}}{\mathbf{res}(i_{t-1}^p)}$;
  - compute $\tilde{w}_{t-1}^p = \frac{w_{t-1}^p(\theta)}{\sum_{\tilde{p}} w_{t-1}^p}$;
  - sample $x_t^p \sim \mathbf{prop}(x_t|\tilde{x}_{t-1}^p)$;
  - compute $w_t^p \leftarrow \tilde{w}_{t-1}^p \frac{p(x_t^p|\tilde{x}_{t-1}^p,\theta)p(y_t|x_t^p,\theta)}{\mathbf{prop}(x_t^p|\tilde{x}_{t-1}^p)}$.

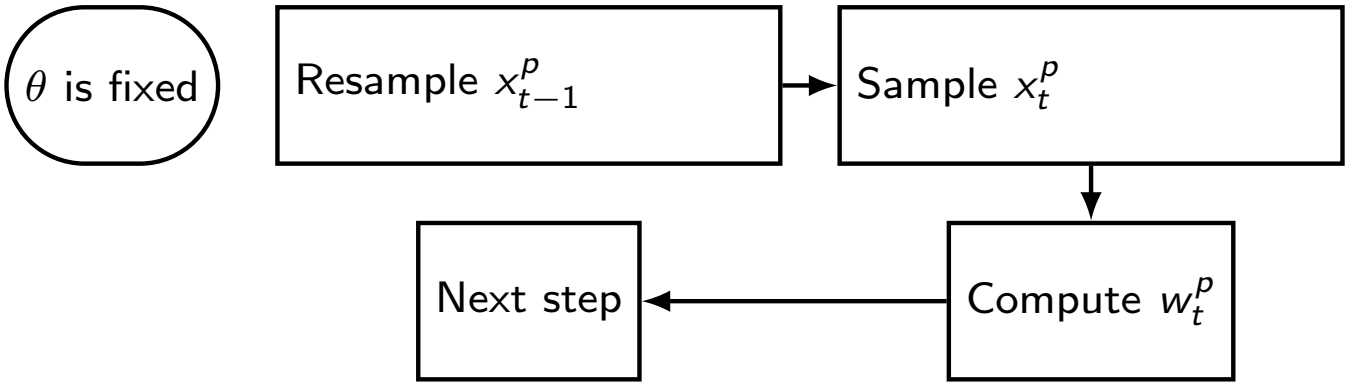
The likelihood is then approximated by:

$$p(y_{1:T}|\theta) \approx \prod_{s=1}^{T} \sum_{p} w_t^p.$$

SMC summary


θ is fixed
Resample x_{t-1}^p
Sample x_t^p
Compute w_t^p
Next step

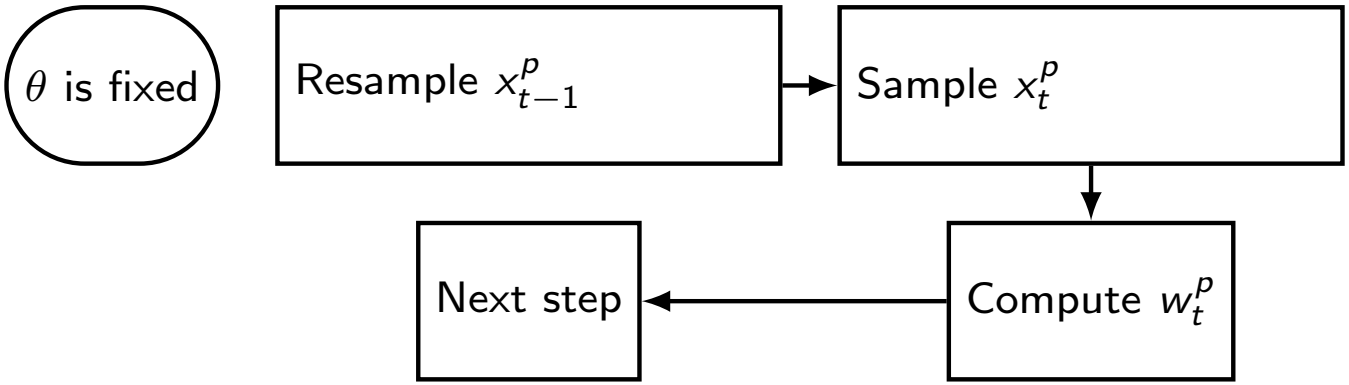
SMC summary



Good news ✓

- We have a good approximation of the likelihood.
- If $\mathbf{prop}(x_t|x_{t-1}) = p(x_t|x_{t-1}, y_{t:T})$ and $\mathbf{res}(i) \propto p(y_{t:T}|x_{t-1}^i)$ it performs incredibly well [Whiteley and Lee, 2014, Rimella et al., 2023a].

SMC summary



Good news ✓

- We have a good approximation of the likelihood.
- If $\textbf{prop}(x_t|x_{t-1}) = p(x_t|x_{t-1}, y_{t:T})$ and $\textbf{res}(i) \propto p(y_{t:T}|x_{t-1}^i)$ it performs incredibly well [Whiteley and Lee, 2014, Rimella et al., 2023a].

Danger ⚠

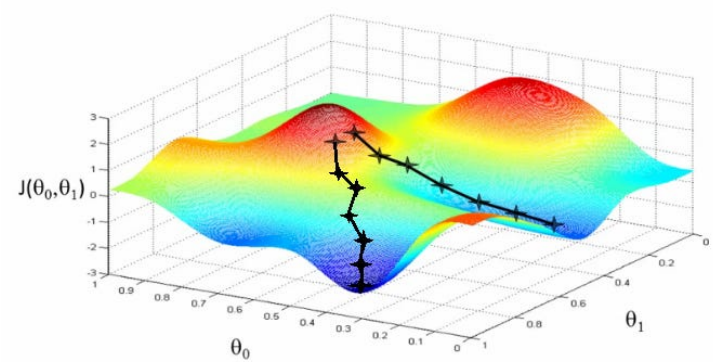
- Proposal distribution and resampling scheme are hard to define, especially in the context of epidemiological modelling.

Parameters inference

- Frequentist: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use any numerical optimizer to find the maximum $\hat{\theta}$.
- Bayesian: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use Markov Chain Monte Carlo (MCMC) to sample from $p(\theta|y_{1:T})$ given a prior $p(\theta)$ and a proposal $\mathbf{q}(\theta|\theta')$:
  - $\theta_{s-1} \sim p(\theta)$
  - Iterate:
    - $\theta_s \sim \mathbf{q}(\theta|\theta_{s-1})$
    - accept θ_s with probability $\frac{p(y_{1:T}|\theta_s)p(\theta_s)\mathbf{q}(\theta_{s-1}|\theta_s)}{p(y_{1:T}|\theta_{s-1})p(\theta_{s-1})\mathbf{q}(\theta_s|\theta_{s-1})}$ otherwise $\theta_s = \theta_{s-1}$

Parameters inference

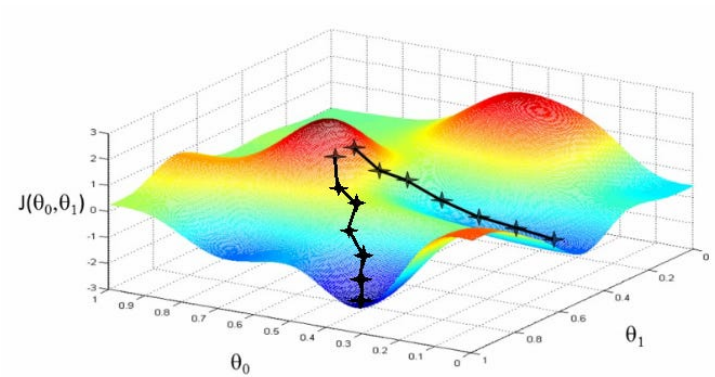
- Frequentist: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use any numerical optimizer to find the maximum $\hat{\theta}$.
- Bayesian: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use Markov Chain Monte Carlo (MCMC) to sample from $p(\theta|y_{1:T})$ given a prior $p(\theta)$ and a proposal $\mathbf{q}(\theta|\theta')$:
  - $\theta_{s-1} \sim p(\theta)$
  - Iterate:
    - $\theta_s \sim \mathbf{q}(\theta|\theta_{s-1})$
    - accept θ_s with probability $\frac{p(y_{1:T}|\theta_s)p(\theta_s)\mathbf{q}(\theta_{s-1}|\theta_s)}{p(y_{1:T}|\theta_{s-1})p(\theta_{s-1})\mathbf{q}(\theta_s|\theta_{s-1})}$ otherwise $\theta_s = \theta_{s-1}$

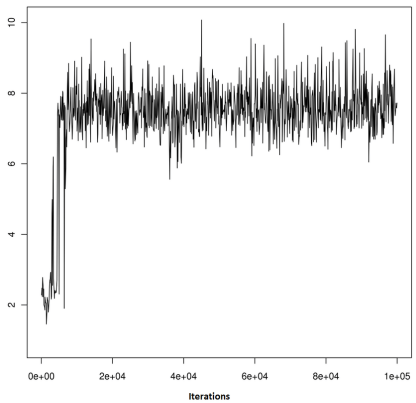


Parameters inference

- Frequentist: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use any numerical optimizer to find the maximum $\hat{\theta}$.
- Bayesian: given $p(y_{1:T}|\theta)$, or an estimate of it, we can use Markov Chain Monte Carlo (MCMC) to sample from $p(\theta|y_{1:T})$ given a prior $p(\theta)$ and a proposal $\mathbf{q}(\theta|\theta')$:
  - $\theta_{s-1} \sim p(\theta)$
  - Iterate:
    - $\theta_s \sim \mathbf{q}(\theta|\theta_{s-1})$
    - accept θ_s with probability $\frac{p(y_{1:T}|\theta_s)p(\theta_s)\mathbf{q}(\theta_{s-1}|\theta_s)}{p(y_{1:T}|\theta_{s-1})p(\theta_{s-1})\mathbf{q}(\theta_s|\theta_{s-1})}$ otherwise $\theta_s = \theta_{s-1}$

  if we do not have direct access to $p(y_{1:T}|\theta_{s-1})$ MCMC is super expensive.





Approximate Bayesian Computation (ABC)

ABC [Kypraios et al., 2017] approximates the posterior $p(\theta|y_{1:t})$:

- repeat for a fixed number of times;
- sample from a prior $\theta^i \sim p(\theta)$;
- sample the data from the model $y^i_{1:t} \sim p(y_{1:t}|\theta^i)$;
- accept θ^i if $y^i_{1:t} \approx y_{1:t}$.

Approximate Bayesian Computation (ABC)

ABC [Kypraios et al., 2017] approximates the posterior $p(\theta|y_{1:t})$:

- repeat for a fixed number of times;
- sample from a prior $\theta^i \sim p(\theta)$;
- sample the data from the model $y^i_{1:t} \sim p(y_{1:t}|\theta^i)$;
- accept θ^i if $y^i_{1:t} \approx y_{1:t}$.

Good news ✓

- We just need to sample from the model.
- We can design even more complicated ABC to tackle complex scenarios.

Approximate Bayesian Computation (ABC)

ABC [Kypraios et al., 2017] approximates the posterior $p(\theta|y_{1:t})$:

- repeat for a fixed number of times;
- sample from a prior $\theta^i \sim p(\theta)$;
- sample the data from the model $y^i_{1:t} \sim p(y_{1:t}|\theta^i)$;
- accept θ^i if $y^i_{1:t} \approx y_{1:t}$.

Good news ✓

- We just need to sample from the model.
- We can design even more complicated ABC to tackle complex scenarios.

Danger ⚠

- We need to define $\approx$, which might affect inference.
- In high-dimensional scenarios it performs poorly.

Practical implementation of SMC and black-box optimizers, and ABC

Exact likelihood

Remark that we can recursively compute the likelihood:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Exact likelihood

Remark that we can recursively compute the likelihood:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Danger ⚠

- The integrals can be computed only in some specific scenarios.

Exact likelihood

Remark that we can recursively compute the likelihood:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \int p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)dx_{t-1}$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
$$p(y_t|y_{1:t-1},\theta) = \int p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)dx_t.$$

Danger ⚠

- The integrals can be computed only in some specific scenarios.

Good news ✓

- The integrals can be computed only in some specific scenarios.

Kalman-Filter [Welch et al., 1995]

Suppose that our state-space model looks like this:

$$x_t = kx_{t-1} + \eta_t,$$
$$y_t = qx_t + \xi_t,$$

where:

$$\eta_t \sim N(0, (\sigma^\eta)^2),$$
$$\xi_t \sim N(0, (\sigma^\xi)^2).$$

Then $x_{0:t}, y_{1:t}$ are jointly Gaussian provided that x_0 is Gaussian.

Kalman-Filter recursion

- "Prediction" given $x_{t-1}|y_{1:t-1} \sim N(m_{t-1}, s^2_{t-1})$:

$$x_t|y_{1:t-1} \sim N(m_{t|t-1}, s^2_{t|t-1}),$$

with $m_{t|t-1} = km_{t-1}$ and $s^2_{t|t-1} = k^2 s^2_{t-1} + (\sigma^\eta)^2$.

- "Update" given $x_t|y_{1:t-1} \sim N(m_{t|t-1}, s^2_{t|t-1})$:

$$[x_t, y_t]^\top|y_{1:t-1} \sim N\left(\begin{pmatrix} m_{t|t-1} \\ M^y_t \end{pmatrix}, \begin{pmatrix} s^2_{t|t-1} & qs^2_{t|t-1} \\ qs^2_{t|t-1} & (s^y_t)^2 \end{pmatrix}\right),$$

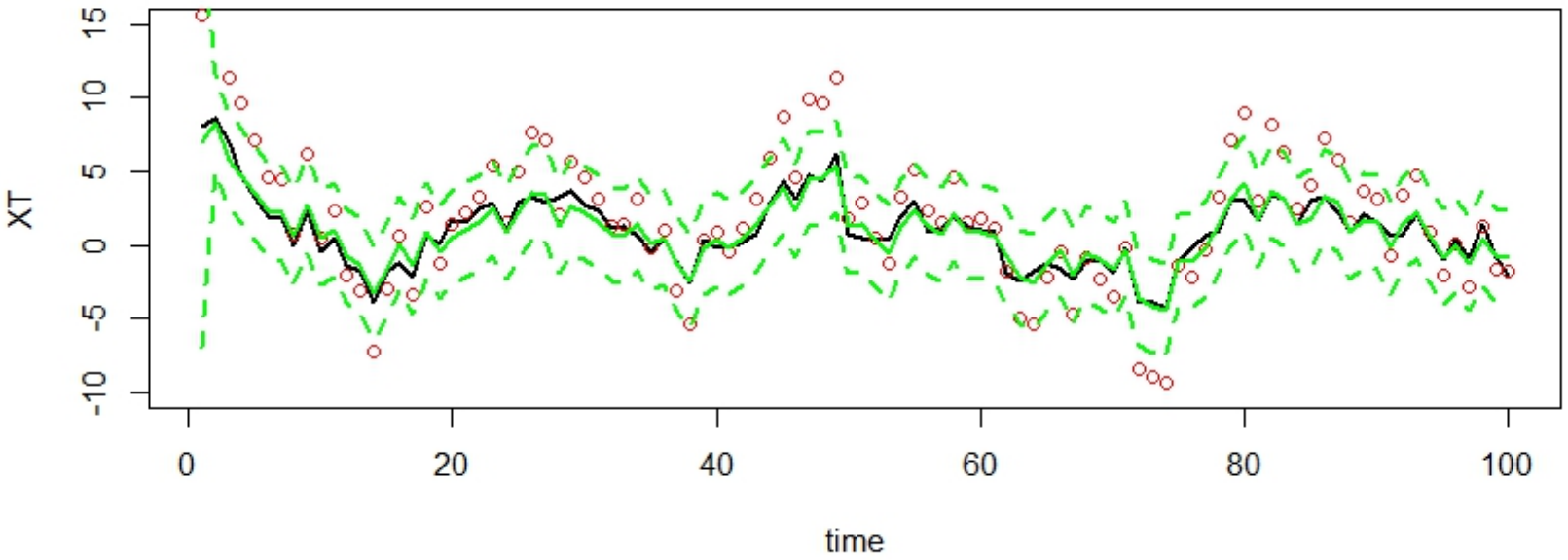
with $M^y_t = qm_{t|t-1}$ and $(s^y_t)^2 = q^2 s^2_{t|t-1} + (\sigma^\xi)^2$:

$$x_t|y_{1:t} \sim N\left(m_{t|t-1} + \frac{qs^2_{t|t-1}(y_t - qm_{t|t-1})}{(s^y_t)^2}, (s^y_t)^2 - \frac{(qs^2_{t|t-1})^2}{(s^y_t)^2}\right),$$

$$y_t|y_{1:t-1} \sim N(qm_{t|t-1}, q^2 s^2_{t|t-1} + (\sigma^\xi)^2).$$

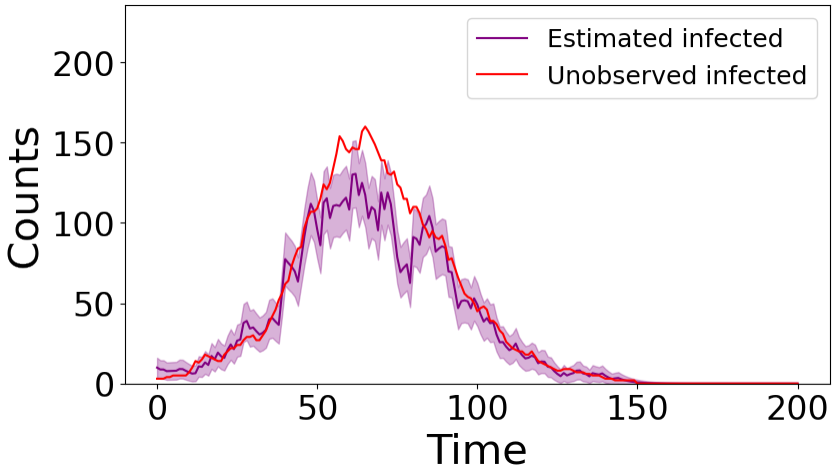
Exploiting conjugacy

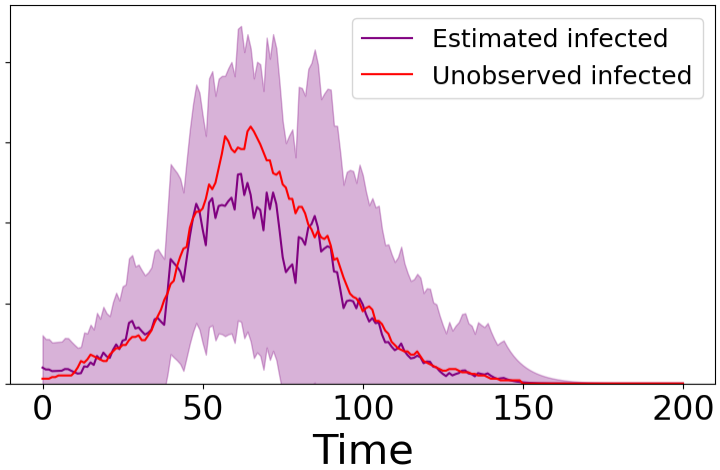
- In the linear Gaussian scenario we have a close form solution for our likelihood.
- All the computations require updating the parameters of a Gaussian.
- The extended-Kalman Filter allows to relax the linearity, but it still needs the noise to be Gaussian [Duffield et al., 2023].



Exploiting conjugacy

- In the linear Gaussian scenario we have a close form solution for our likelihood.
- All the computations require updating the parameters of a Gaussian.
- The extended-Kalman Filter allows to relax the linearity, but it still needs the noise to be Gaussian [Duffield et al., 2023].
- Can we find other class of distributions with similar properties?





General compartmental model

Remember the general formulation of compartmental models:

- the dynamics is:

$$x_0 \sim \textbf{Multi}(N, p_0)$$

$$x_t^{(i)} = \sum_{j=1}^{M} Z_{t-1,t}^{(j,i)} \text{ for } i = 1, \ldots, M$$

$$\text{with } Z_{t-1,t}^{(i,\cdot)} \sim \textbf{Multi}(x_{t-1}^{(i)}, K_{x_{t-1}}^{(i,\cdot)}) \text{ for } i = 1, \ldots, M;$$

- the observations are:

$$y_t^{(i)} \sim \textbf{Bin}(x_t^{(i)}, q^{(i)}), \text{ with } i = 1, \ldots, M.$$

PAL/MAL [Whitehouse et al., 2023, Whiteley and Rimella, 2021]

- What?
  We want to approximate the filtering distribution and likelihood, via a parametric distribution modelling the counts of N individuals over M compartments.

- How?
  We can just modify "Prediction" and "Update" to keep the approximation in the same class.

PAL/MAL [Whitehouse et al., 2023, Whiteley and Rimella, 2021]

- What?
  We want to approximate the filtering distribution and likelihood, via a parametric distribution modelling the counts of N individuals over M compartments.
  $\longrightarrow \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
- How?
  We can just modify "Prediction" and "Update" to keep the approximation in the same class.

PAL/MAL [Whitehouse et al., 2023, Whiteley and Rimella, 2021]

- What?
  We want to approximate the filtering distribution and likelihood, via a parametric distribution modelling the counts of N individuals over M compartments.
  $\longrightarrow \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
- How?
  We can just modify "Prediction" and "Update" to keep the approximation in the same class.
  $\longrightarrow x_t|y_{1:t-1} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
  $\longrightarrow x_t|y_{1:t} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$

PAL/MAL [Whitehouse et al., 2023, Whiteley and Rimella, 2021]

- What?
  We want to approximate the filtering distribution and likelihood, via a parametric distribution modelling the counts of N individuals over M compartments.
  $\longrightarrow \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
- How?
  We can just modify "Prediction" and "Update" to keep the approximation in the same class.
  $\longrightarrow x_t|y_{1:t-1} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
  $\longrightarrow x_t|y_{1:t} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\cdot)$ or $\textbf{Multi}(N, \cdot)$
- Why?
  Orders of magnitude faster and good quality approximation [Whiteley and Rimella, 2021, Whitehouse et al., 2023].

Exact likelihood

We can recursively compute the likelihood:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \sum_{x_{t-1}} p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
   $p(y_t|y_{1:t-1},\theta) = \sum_{x_t} p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)$.

$\longrightarrow$ Compute the marginal likelihood $p(y_1|\theta)\prod_{s=2}^{t} p(y_t|y_{1:s-1},\theta)$.
$\longrightarrow$ Estimate the static parameters.

Approximate "Prediction"

- Given $x_{t-1}|y_{1:t-1} \sim \textbf{Multi}(N, \pi_{t-1})$ or $x_{t-1}|y_{1:t-1} \sim \bigotimes_{i=1}^{M} \textbf{Pois}(\lambda_{t-1}^{(i)})$.
- Use $K_{N\pi_{t-1}}$ or $K_{\lambda_{t-1}}$ in the prediction step.
- Get:

$$x_t|y_{1:t-1} \approx \textbf{Multi}(N, \pi_{t|t-1}) \text{ with } \pi_{t|t-1} = (\pi_{t-1}^\top K_{N\pi_{t-1}})^\top \text{ or}$$

$$x_t|y_{1:t-1} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\lambda_{t|t-1}^{(i)}) \text{ with } \lambda_{t|t-1} = (\lambda_{t-1}^\top K_{\lambda_{t-1}})^\top.$$

Good news ✓

We just require operations on M dimensional vectors and matrices.

Forward step and marginal likelihood

We can recursively compute the likelihood:

1. Initialization: $p(x_0|\theta)$ is given;
2. "Prediction": $p(x_t|y_{1:t-1},\theta) = \sum_{x_{t-1}} p(x_t|x_{t-1},\theta)p(x_{t-1}|y_{1:t-1},\theta)$;
3. "Update": $p(x_t|y_{1:t},\theta) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1},\theta)}{p(y_t|y_{1:t-1},\theta)}$,
   $p(y_t|y_{1:t-1},\theta) = \sum_{x_t} p(y_t|x_t,\theta)p(x_t|y_{1:t-1},\theta)$.

$\longrightarrow$ Compute the marginal likelihood $p(y_1|\theta)\prod_{s=2}^{t} p(y_t|y_{1:s-1},\theta)$.
$\longrightarrow$ Estimate the static parameters.

Approximate "Update"

- Given $x_t|y_{1:t-1} \sim \textbf{Multi}(N, \pi_{t|t-1})$ or $x_t|y_{1:t-1} \sim \bigotimes_{i=1}^{M} \textbf{Pois}(\lambda_{t|t-1}^{(i)})$.
- Get:

$$x_t|y_{1:t-1} \approx \textbf{Multi}(N, \pi_t) \text{ with } \pi_t = \frac{y_t}{N} + \left(1 - \frac{\mathbf{1}_m^\top y_t}{N}\right) \frac{\pi_{t|t-1} \circ (\mathbf{1}_m - q)}{1 - \pi_{t|t-1}^\top q} \text{ or}$$

$$x_t|y_{1:t-1} \approx \bigotimes_{i=1}^{M} \textbf{Pois}(\lambda_t^{(i)}) \text{ with } \lambda_t = y_t + \lambda_{t|t-1} \circ (\mathbf{1}_m - q).$$

- With also:

$$\begin{pmatrix} y_t \\ N - \mathbf{1}_m^\top y_t \end{pmatrix} |y_{1:t-1} \sim \textbf{Multi}\left(N, \begin{pmatrix} \pi_{t|t-1} \circ q \\ 1 - \pi_{t|t-1}^\top q \end{pmatrix}\right) \text{ or } y_t \sim \bigotimes_{i=1}^{M} \textbf{Pois}(q^{(i)} \lambda_{t|t-1}^{(i)})$$

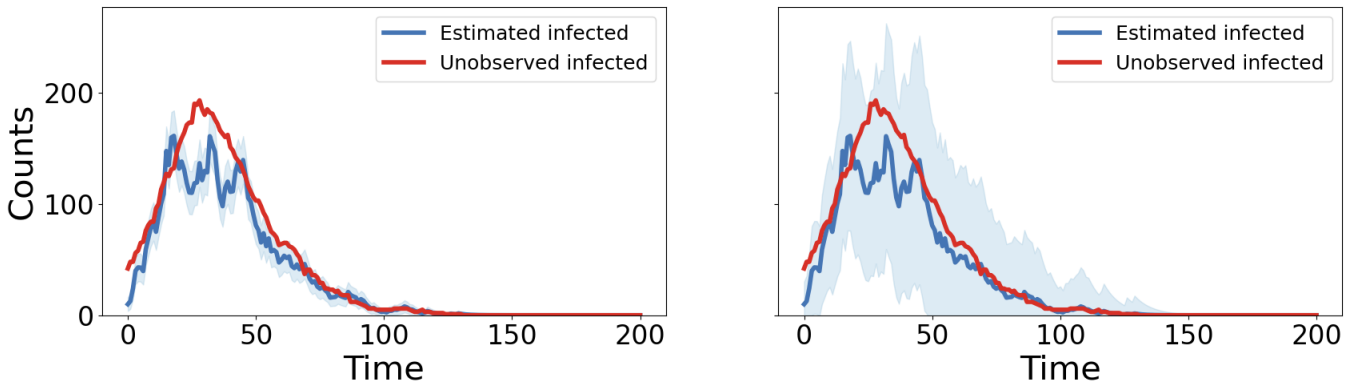
Good news ✓

- We just require operations on M dimensional vectors and matrices.
- It is partially approximate cause the likelihood is exact.

Overdispersion

Sometimes the data are very noisy...

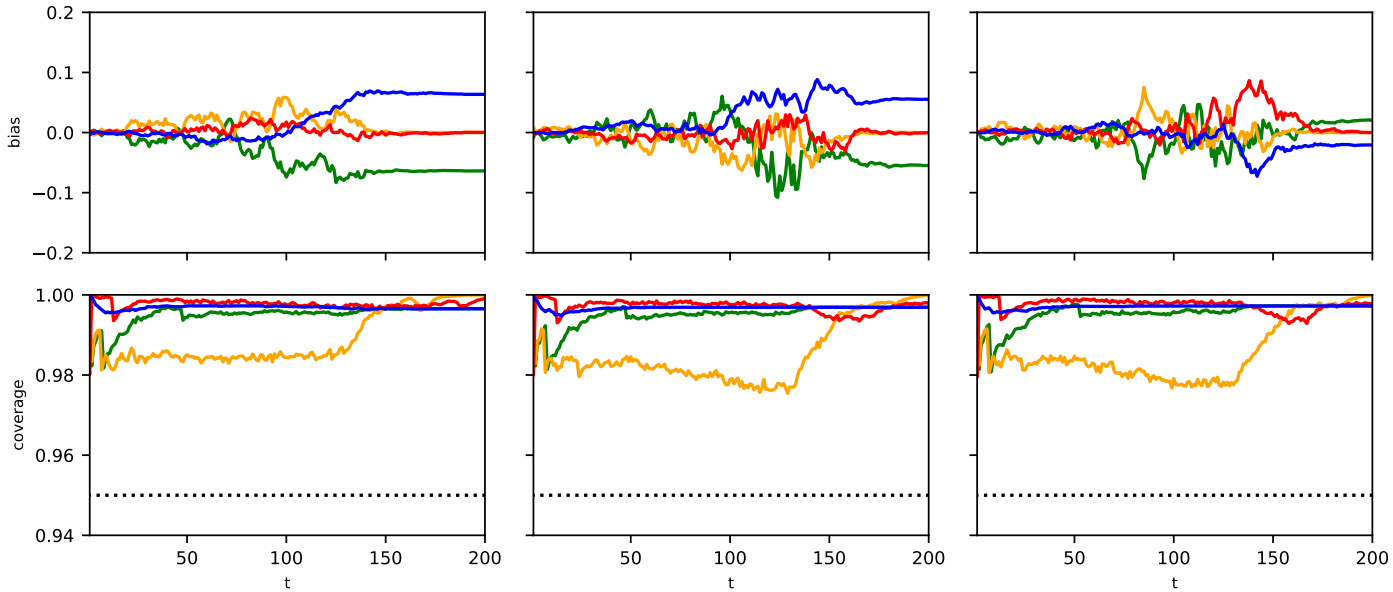
- From a modelling perspective we can include extra noise by using compound distributions, e.g. use a Beta-Binomial observation model and a stochastic infection rate.
- This makes inference way more complicated.
  - We can nest the PAL/MAL in an SMC.
  - The MAL/PAL will take care of the compartments states while the SMC will take care of the noise in the parameters.



Bias and credible interval coverage

In the Ebola model choose a combination of parameters θ, an initial condition π_0 and an increasing population size, i.e. $N = 5 \times 10^2, 5 \times 10^4, 5 \times 10^6$. We simulate 2×10^4 epidemics over 200 time steps.

- Compute the empirical bias between $N\pi_{t|t}$ and x_t.
- Compute the coverage of the 95%-credible interval for the marginal over each $x_t^{(i)}$.

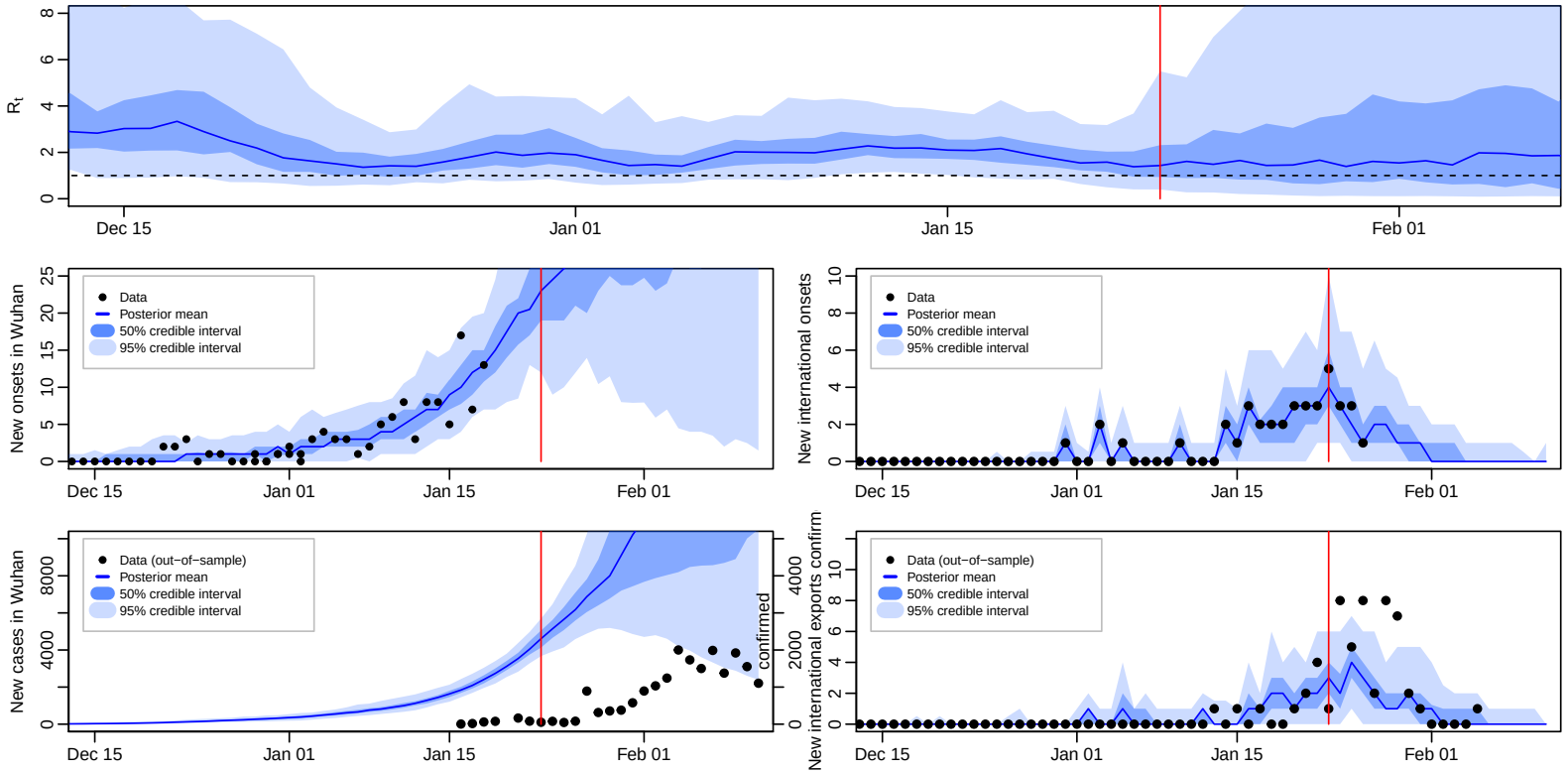


Red, yellow, blue, green correspond to $x_t^{(i)}$, $i = S, E, I, R$.

COVID-19

Estimation of the reproduction number of COVID-19 in Wuhan as in [Kucharski et al., 2020].

- Consider 10 compartments (in Wuhan, outside Wuhan, epidemic stages, ...).
- Consider a time-varying $(\beta_t)_{t \geqslant 0}$ with a-priori by a geometric random walk.
- Embed the parametric approximation in an SMC.



Measles model

Consider epidemics in $J = 40$ cities across the UK over the 16 year with dynamic:

$$S_{k,t+1} = S_{k,t} - B_{k,t} - F^{(S)}_{k,t} + A_{k,t}, \qquad E_{k,t+1} = E_{k,t} + B_{k,t} - C_{k,t} - F^{(E)}_{k,t},$$
$$I_{k,t+1} = I_{k,t} + C_{k,t} - D_{k,t} - F^{(I)}_{k,t}, \qquad R_{k,t+1} = R_{k,t} + D_{k,t} - F^{(R)}_{k,t},$$

where $F^{(\cdot)}_{t,k}$ are the deaths and $A_{k,t}$ are the births; and $C_{k,t}$ and $D_{k,t}$ are binomially distributed. The term $B_{k,t}$ is

$$B_{k,t} \sim \text{Bin}\left(S_{k,t} - F^{(S)}_{k,t}, 1 - e^{-hb_{k,r}}\right),$$

where for $r \geqslant 1$, $t = \tau_r, \ldots, \tau_{r+1} - 1$:

$$b_{k,r} = \beta_{k,r}\xi_{k,r} \cdot \left[\left(\frac{I_{k,\tau_r}}{n_{k,\tau_r}}\right) + \sum_{l \neq k} \frac{v_{k,l}}{n_{k,\tau_r}} \left\{\left(\frac{I_{l,\tau_r}}{n_{l,\tau_r}}\right) - \left(\frac{I_{k,\tau_r}}{n_{k,\tau_r}}\right)\right\}\right].$$

Model	No. parameters	Log-likelihood;(sd)	AIC	Comp. time
A	11	−63579; (62)	127180	45 min
B	128	−61257; (28)	122770	10 hr
C	167	−61169; (34)	122672	24 hr
Park and Ion.,2020	12	−70000	140024	30 hr

Practical implementation of the PAL and gradient ascent

SIR from another perspective

Let ι_t^n be the compartment of individual n at time t then x_t is given by:

$$x_t^{(i)} = \sum_{n=1}^{N} \mathbb{I}[\iota_t^n = i], \quad \text{for } i = S, I, R.$$

For each $n = 1, \ldots, N$:

$$\iota_t^n \Big| \left(\iota_{t-1}^n\right)_{n=1,\ldots,N} \sim K_{x_{t-1}}^{(\iota_{t-1}^n, \cdot)}$$

where:

$$K_x = \begin{pmatrix} e^{-\beta \frac{x^{(I)}}{N}} & 1 - e^{-\beta \frac{x^{(I)}}{N}} & 0 \\ 0 & e^{-\gamma} & 1 - e^{-\gamma} \\ 0 & 0 & 1 \end{pmatrix}.$$

K_x is the individuals' kernel and depends on the current state of the population.

IBM Motivation

Inference in individual-based models (IBM) is as interesting as complicated:

- IBM is the closest to reality, all individuals are different and they have different interactions;
- we have detailed information on the individuals;

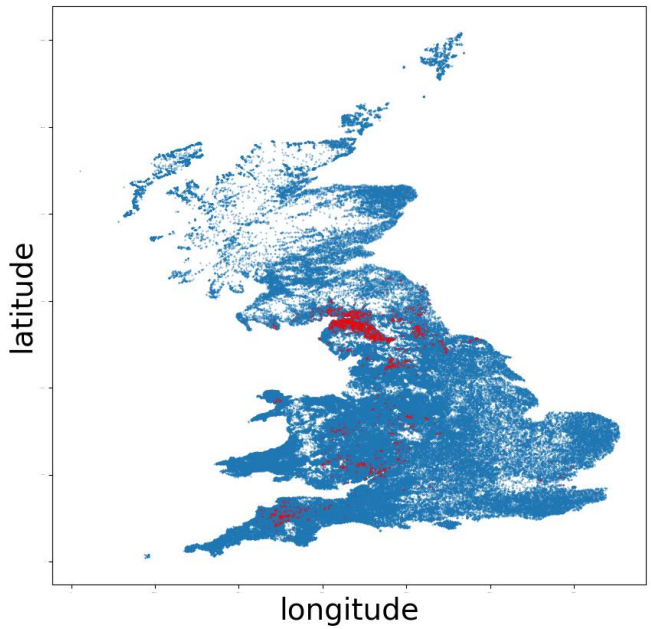
but:

- close-form inference is exponential in the number of individuals;
- SMC algorithm fails without smart proposal distribution, which are expensive;
- other simulation based algorithms (e.g. ABC) easily fail for the dimensionality of the model.

Definition

Notation

- M is the number of compartments;
- N is the population size;
- $(\mathbf{p}_{n,0})_{n\in[1:N]}$ is the initial distribution;
- $(\mathbf{K}_{n,\cdot})_{n\in[1:N]}$ is the transition kernel;



Dynamics

We use $(\iota_t)_{t\geqslant 0}$ for the process indicating the population state:

Time 0: $\iota_0^n \sim \mathbf{Cat}\,(\mathbf{p}_{n,0})$ for $n \in [1:N]$;

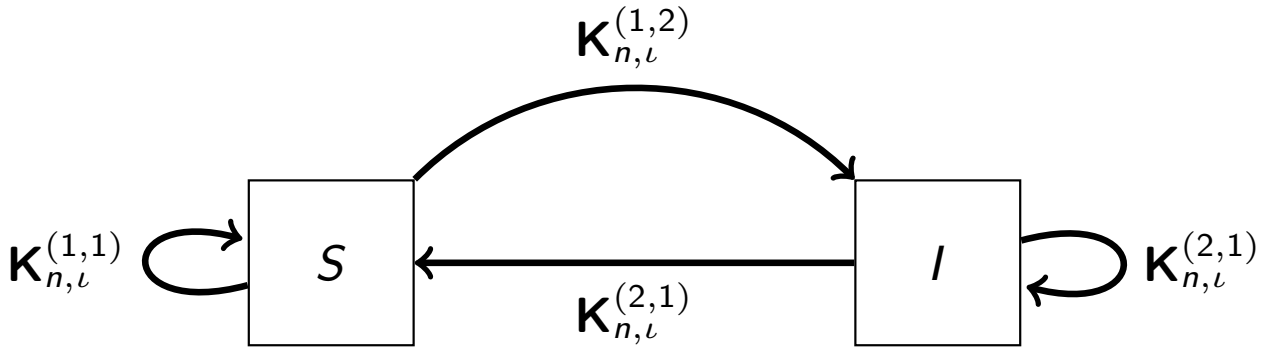
Time t: $\iota_t^n | \iota_{t-1} \sim \mathbf{Cat}\left(\mathbf{K}_{n,\iota_{t-1}}^{(\iota_{t-1}^n,\cdot)}\right)$ for $n \in [1:N]$.

A simple SIS example from [Ju et al., 2021]

Consider $d \in \mathbb{N}$ covariates for each individual gathered in $(\mathbf{w}_n)_{n\in[1:N]}$, then we can compute for $n \in [1:N]$:

$$\mathbf{p}_{n,0} = \begin{bmatrix} 1 - \frac{1}{1+\exp\left(-\beta_0^{\mathrm{T}} w_n\right)} \\ \frac{1}{1+\exp\left(-\beta_0^{\mathrm{T}} w_n\right)} \end{bmatrix}, \quad \mathbf{K}_{n,\iota_t} = \begin{bmatrix} 1 - \frac{1}{1+\exp\left(-\beta_\lambda^{\mathrm{T}} w_n\right)} \frac{x_t^{(I)}}{N} & \frac{1}{1+\exp\left(-\beta_\lambda^{\mathrm{T}} w_n\right)} \frac{x_t^{(I)}}{N} \\ \frac{1}{1+\exp\left(-\beta_\gamma^{\mathrm{T}} w_n\right)} & 1 - \frac{1}{1+\exp\left(-\beta_\gamma^{\mathrm{T}} w_n\right)} \end{bmatrix}$$

with $x_t^{(I)} := \sum_{n\in[1:N]} \mathbb{I}\left(\iota_t^n = I\right)$ and $\beta_0 \in \mathbb{R}^d$ and $\beta_\lambda, \beta_\gamma \in \mathbb{R}^d$. In this model we have individual-specific probabilities of infection and recovery.



Likelihood

The observations are denoted by $(y_t)_{t\geqslant 1}$ and given $(\mathbf{q}_{n,t})_{n\in[1:N],t\geqslant 1}$ with $\mathbf{q}_{n,t} \in [0,1]^M$ per each time step t:

$$y_t^n = \iota_t^n \mathbf{r}_t^n \text{ with } \mathbf{r}_t^n \sim \mathbf{Be}\left(\cdot \middle| \mathbf{q}_{n,t}^{(\iota_t^n)}\right) \text{ for } n \in [1:N]. \tag{1}$$

The likelihood is:

$$p(y_{1:t}|\theta) := \sum_{\iota_{0:t}} p(y_{1:t}|\iota_{0:t},\theta)p(\iota_{0:t}|\theta)$$

remark that θ is used to denote the entire collection of parameters in the model.

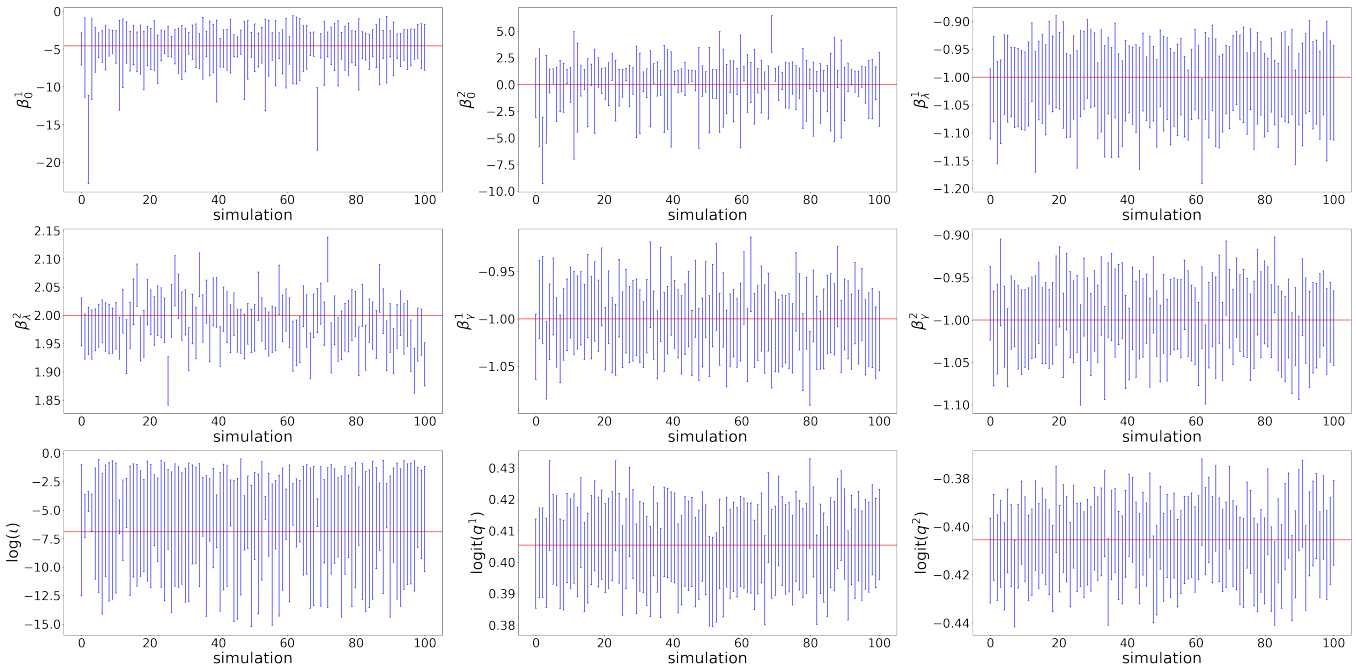
Danger ⚠

The likelihood computation requires marginalizing out recursively ι_{t-1} and ι_t, i.e. we have a computational cost of $\mathcal{O}(TM^{2N})$!

Possible strategies

Some approaches can be followed to compute the likelihood:

- define smart proposal distribution in SMC by looking into future observations [Rimella et al., 2023a];
- condition on other individuals via Monte Carlo simulation [Rimella et al., 2023b].



WHAT IS WRONG WITH SIR?

During Bayescomp2023 Simon Spencer asked the audience:

"WHAT IS WRONG WITH SIR?"

WHAT IS WRONG WITH SIR?

- No exposed period
- No asymptomatics
- No symptomatic period
- No reinfections
- No demography
- No households
- No schools
- No social network structure
- No spatial structure
- No age structure
- No sexual contact structure
- No vaccines
- No treatments
- No testing
- No hospitalisation
- No healthcare workers
- No care homes
- No immunocompromised
- No variants
- No pathogen genetics
- No pathogen evolution
- No immunology
- No behaviour change
- No vectors
- No animal reservoir
- No seasonality
- No infectivity profiles
- No lockdown

Conclusion

We can be as creative as we want in our model the important things are that it matches the biology of the pathogen and that we are able to fit it :)

Thank you!

References I

Chopin, N. and Papaspiliopoulos, O. (2020).
Introduction to Sequential Monte Carlo.
Springer International Publishing.

Duffield, S., Power, S., and Rimella, L. (2023).
A state-space perspective on modelling and inference for online skill rating.
arXiv preprint arXiv:2308.02414.

Ju, N., Heng, J., and Jacob, P. E. (2021).
Sequential monte carlo algorithms for agent-based models of disease transmission.
arXiv preprint arXiv:2101.12156.

Keeling, M. J. and Rohani, P. (2011).
Modeling infectious diseases in humans and animals.
In *Modeling infectious diseases in humans and animals*. Princeton university press.

References II


Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R. M., Sun, F., Jit, M., Munday, J. D., et al. (2020).
Early dynamics of transmission and control of covid-19: a mathematical modelling study.
The lancet infectious diseases, 20(5):553–558.


Kypraios, T., Neal, P., and Prangle, D. (2017).
A tutorial introduction to bayesian inference for stochastic epidemic models using approximate bayesian computation.
Mathematical biosciences, 287:42–53.


Rimella, L., Jewell, C., and Fearnhead, P. (2023a).
Approximating Optimal SMC Proposal Distributions in Individual-Based Epidemic Models.
Statistica Sinica, pages SS–2022–0198.


Rimella, L., Jewell, C., and Fearnhead, P. (2023b).
Simulation based composite likelihood.
arXiv preprint arXiv:2310.10761.

References III


Welch, G., Bishop, G., et al. (1995).
An introduction to the kalman filter.


Whitehouse, M., Whiteley, N., and Rimella, L. (2023).
Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods.
Journal of the Royal Statistical Society Series B: Statistical Methodology, page qkad065.


Whiteley, N. and Lee, A. (2014).
Twisted particle filters.
The Annals of Statistics.


Whiteley, N. and Rimella, L. (2021).
Inference in stochastic epidemic models via multinomial approximations.
In *International Conference on Artificial Intelligence and Statistics*, pages 1297–1305. PMLR.