Consistent and fast inference in compartmental models of epidemics via Poisson Approximate Likelihoods.

Michael Whitehouse
Nick Whiteley
Lorenzo Rimella

University of Bristol

2022

Talk outline

1. Compartmental models
   - Simple *SEIR* example
   - Latent Compartmental Model
   - Likelihood Intractibility
2. Poisson Approximate Likelihoods
   - The approximation
3. Consistency
   - Set up
   - Theorem
4. Concluding remarks

SEIR model

Consider a population size n and let $\mathrm{x}_t = [\mathrm{x}_t^{(S)}, \mathrm{x}_t^{(E)}, \mathrm{x}_t^{(I)}, \mathrm{x}_t^{(R)}]^{\mathrm{T}}$:

$$
\begin{aligned}
\mathrm{x}_{t+1}^{(S)} &= \mathrm{x}_t^{(S)} - B_t,\\
\mathrm{x}_{t+1}^{(E)} &= \mathrm{x}_t^{(E)} + B_t - C_t,\\
\mathrm{x}_{t+1}^{(I)} &= \mathrm{x}_t^{(I)} + C_t - D_t,\\
\mathrm{x}_{t+1}^{(R)} &= \mathrm{x}_t^{(R)} + D_t,
\end{aligned}
$$

with

$$
\begin{aligned}
B_t &\sim \mathrm{Bin}\left(\mathrm{x}_t^{(S)}, 1 - e^{-\beta \mathrm{x}_t^{(I)}/n}\right),\\
C_t &\sim \mathrm{Bin}\left(\mathrm{x}_t^{(E)}, 1 - e^{-\rho}\right),\\
D_t &\sim \mathrm{Bin}\left(\mathrm{x}_t^{(I)}, 1 - e^{-\gamma}\right).
\end{aligned}
$$

General Latent Compartmental Model

(Whiteley and Rimella, 2021).

Let $\xi_t^{(k)}$ be the location of individual k at time t then x_t is given by:

$$\mathrm{x}_t^{(i)} = \sum_{k=1}^{n} \mathbb{I}[\xi_t^{(k)} = i], \quad \text{for } i = S, E, I, R.$$

For each $k = 1, \dots, n$:

$$\xi_0^{(k)} \sim \pi_0 \quad \text{and} \quad \xi_t^{(k)} \Big| \left(\xi_{t-1}^{(k)}\right)_{k=1,\dots,n} \sim K_{t,\boldsymbol{\eta}(\mathrm{x}_{t-1})}^{(\xi_{t-1}^{(k)},\cdot)}$$

with $\eta^{(i)}(\mathrm{x}_t) = \mathrm{x}_t^{(i)}/n$ for $i \in \{S, E, I, R\}$ and:

$$K_{t,\boldsymbol{\eta}(\mathrm{x})} = \begin{pmatrix} e^{-\beta\eta^{(I)}(\mathrm{x})} & 1 - e^{-\beta\eta^{(I)}(\mathrm{x})} & 0 & 0 \\ 0 & e^{-\rho} & 1 - e^{-\rho} & 0 \\ 0 & 0 & e^{-\gamma} & 1 - e^{-\gamma} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

General Latent Compartmental Model

(Whiteley and Rimella, 2021).

Let $\xi_t^{(k)}$ be the location of individual k at time t then x_t is given by:

$$\mathrm{x}_t^{(i)} = \sum_{k=1}^{n} \mathbb{I}[\xi_t^{(k)} = i], \quad \text{for } i = S, E, I, R.$$

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

$$\xi_0^{(k)} \sim \pi_0 \quad \text{and} \quad \xi_{\mathrm{t}}^{(\mathrm{k})} \Big| \left(\xi_{\mathrm{t-1}}^{(\mathrm{k})}\right)_{\mathrm{k}=1,\ldots,\mathrm{n}} \sim \mathrm{K}_{\mathrm{t},\boldsymbol{\eta}(\mathrm{x}_{\mathrm{t-1}})}^{(\xi_{\mathrm{t-1}}^{(\mathrm{k})},\cdot)}$$

with $\eta^{(i)}(\mathrm{x}_t) = \mathrm{x}_t^{(i)}/n$ for $i \in \{S, E, I, R\}$ and:

$$\mathrm{K}_{t,\boldsymbol{\eta}(\mathrm{x})} = \begin{pmatrix} e^{-\beta\eta^{(I)}(\mathrm{x})} & 1 - e^{-\beta\eta^{(I)}(\mathrm{x})} & 0 & 0 \\ 0 & e^{-\rho} & 1 - e^{-\rho} & 0 \\ 0 & 0 & e^{-\gamma} & 1 - e^{-\gamma} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

We do not observe the full compartments x_t, indeed for $q_t = [q_t^{(S)}, q_t^{(E)}, q_t^{(I)}, q_t^{(R)}]^{\mathrm{T}}$:

$$y_t^{(i)} \sim \mathrm{Bin}(x_t^{(i)}, q_t^{(i)}), \text{ with } i \in \{S, E, I, R\}.$$

Generalization to m compartments

Let $\xi_t^{(k)}$ be the compartment of individual k at time t then x_t is given by:

$$\mathrm{x}_t^{(i)} = \sum_{k=1}^{n} \mathbb{I}[\xi_t^{(k)} = i], \quad \text{for } i = 1, \ldots, m.$$

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

$$\xi_0^{(k)} \sim \pi_0 \quad \text{and} \quad \xi_t^{(k)} \Big| \left(\xi_{t-1}^{(k)}\right)_{k=1,\ldots,n} \sim K_{t,\boldsymbol{\eta}(\mathrm{x}_{t-1})}^{(\xi_{t-1}^{(k)},\cdot)}$$

with $\eta^{(i)}(\mathrm{x}_t) = \mathrm{x}_t^{(i)}/n$ for $i \in \{1, \ldots, m\}$ and:

$K_{t,\boldsymbol{\eta}(\mathrm{x})}$ is a stochastic matrix with m rows and m columns.

We do not observe the full compartments x_t, indeed for $q_t = [q_t^{(1)}, \ldots, q_t^{(m)}]^{\mathrm{T}}$:

$$y_t^{(i)} \sim \mathrm{Bin}(x_t^{(i)}, q_t^{(i)}), \text{ with } i \in \{1, \ldots, m\}.$$

Model extensions

We (Whitehouse et al., 2022) extended the model to include:

- Immigration

Model extensions

We (Whitehouse et al., 2022) extended the model to include:

- Immigration
- Emigration

Model extensions

We (Whitehouse et al., 2022) extended the model to include:

- Immigration
- Emigration
- Mis-reporting

Model extensions

We (Whitehouse et al., 2022) extended the model to include:

- Immigration
- Emigration
- Mis-reporting
- Spurious reporting

Remark:

$(x_t, y_t)_{t \geqslant 0}$ *is a Hidden Markov model (HMM)*

Inference

Given $p(x_0) =: p(x_0 \mid y_{1:0})$

$$p(x_{t-1}|y_{1:t-1}) \xrightarrow{\text{prediction}} p(x_t|y_{1:t-1}) \xrightarrow{\text{update}} p(x_t|y_{1:t}),$$

Inference

Given $p(\mathrm{x}_0) =: p(\mathrm{x}_0 \mid y_{1:0})$

$$p(\mathrm{x}_{t-1}|\mathrm{y}_{1:t-1}) \overset{\text{prediction}}{\longrightarrow} p(\mathrm{x}_t|\mathrm{y}_{1:t-1}) \overset{\text{update}}{\longrightarrow} p(\mathrm{x}_t|\mathrm{y}_{1:t}),$$

where, for $t \geqslant 1$,

$$p(\mathrm{x}_t|\mathrm{y}_{1:t-1}) = \sum_{\mathrm{x}_{t-1}\in\mathbb{N}_0^m} p(\mathrm{x}_t|\mathrm{x}_{t-1})p(\mathrm{x}_{t-1}|\mathrm{y}_{1:t-1}),$$

$$p(\mathrm{x}_t|\mathrm{y}_{1:t}) = \frac{p(\mathrm{y}_t|\mathrm{x}_t)p(\mathrm{x}_t|\mathrm{y}_{1:t-1})}{p(\mathrm{y}_t|\mathrm{y}_{1:t-1})},$$

$$p(\mathrm{y}_t|\mathrm{y}_{1:t-1}) = \sum_{\mathrm{x}_t\in\mathbb{N}_0^m} p(\mathrm{y}_t|\mathrm{x}_t)p(\mathrm{x}_t|\mathrm{y}_{1:t-1}),$$

Inference

Given $p(x_0) =: p(x_0 \mid y_{1:0})$

$$p(x_{t-1}|y_{1:t-1}) \xrightarrow{\text{prediction}} p(x_t|y_{1:t-1}) \xrightarrow{\text{update}} p(x_t|y_{1:t}),$$

where, for $t \geqslant 1$,

$$p(x_t|y_{1:t-1}) = \sum_{x_{t-1}\in\mathbb{N}_0^m} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}),$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})},$$

$$p(y_t|y_{1:t-1}) = \sum_{x_t\in\mathbb{N}_0^m} p(y_t|x_t)p(x_t|y_{1:t-1}),$$

$$p(y_{1:t}) = \prod_{s=1}^{t} p(y_s|y_{1:s-1}).$$

Inference

Given $p(x_0) =: p(x_0 \mid y_{1:0})$

$$p(x_{t-1}|y_{1:t-1}) \xrightarrow{\text{prediction}} p(x_t|y_{1:t-1}) \xrightarrow{\text{update}} p(x_t|y_{1:t}),$$

where, for $t \geqslant 1$,

$$p(x_t|y_{1:t-1}) = \sum_{x_{t-1} \in \mathbb{N}_0^m} p(x_t|x_{t-1}) p(x_{t-1}|y_{1:t-1}),$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t) p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})},$$

$$p(y_t|y_{1:t-1}) = \sum_{x_t \in \mathbb{N}_0^m} p(y_t|x_t) p(x_t|y_{1:t-1}),$$

$$p(y_{1:t}) = \prod_{s=1}^{t} p(y_s|y_{1:s-1}).$$

Circumventing the issue

Circumventing the issue

- Simulation based inference:
    - Sequential Monte Carlo
    - Approximate Bayesian computation

Circumventing the issue

- Simulation based inference:
  - Sequential Monte Carlo
  - Approximate Bayesian computation
- Deterministic Modelling
  - ODEs

Circumventing the issue

- Simulation based inference:
  - Sequential Monte Carlo
  - Approximate Bayesian computation
- Deterministic Modelling
  - ODEs
- Other approximate inference methods
  - Linear Noise Approximation

Inference

Given $p(x_0) =: p(x_0 \mid y_{1:0})$

$$p(x_{t-1}|y_{1:t-1}) \xrightarrow{\text{prediction}} p(x_t|y_{1:t-1}) \xrightarrow{\text{update}} p(x_t|y_{1:t}),$$

where, for $t \geqslant 1$,

$$p(x_t|y_{1:t-1}) = \sum_{x_{t-1}\in\mathbb{N}_0^m} p(x_t|x_{t-1})p(x_{t-1}|y_{1:t-1}),$$

$$p(x_t|y_{1:t}) = \frac{p(y_t|x_t)p(x_t|y_{1:t-1})}{p(y_t|y_{1:t-1})},$$

$$p(y_t|y_{1:t-1}) = \sum_{x_t\in\mathbb{N}_0^m} p(y_t|x_t)p(x_t|y_{1:t-1}),$$

$$p(y_{1:t}) = \prod_{s=1}^{t} p(y_s|y_{1:s-1}).$$

Introducing the PAL

The general idea of the PAL is to obtain vector-Poisson distribution approximation to each of the terms $p(\mathrm{y}_1)$ and $p(\mathrm{y}_t|\mathrm{y}_{1:t-1})$, $t \geqslant 1$, computed via vector-Poisson approximations to each of the filtering distributions, i.e.

$$p(\mathrm{x}_t \mid \mathrm{y}_{1:t-1}) \approx \mathrm{Poi}(\boldsymbol{\lambda}_t),$$

$$p(\mathrm{x}_t \mid \mathrm{y}_{1:t}) \approx \mathrm{Poi}(\bar{\boldsymbol{\lambda}}_t),$$

$$p(\mathrm{y}_t \mid \mathrm{y}_{1:t-1}) \approx \mathrm{Poi}(\boldsymbol{\mu}_t),$$

where $\mathrm{x} \sim \mathrm{Poi}(\boldsymbol{\lambda})$ means $x^{(i)} \sim \mathrm{Poi}(\lambda^{(i)})$.

Approximating the prediction step

For $\mathrm{x} \in \mathbb{R}^m$ and a length-m probability vector $\boldsymbol{\eta}$, let $M_t(\mathrm{x}, \boldsymbol{\eta}, \cdot)$ be the transition kernel induced by $\mathrm{K}_{t,\boldsymbol{\eta}}$. Then we have:

$$
\begin{aligned}
p(\mathrm{x}_t | \mathrm{y}_{1:t-1}) &= \sum_{\mathrm{x}_{t-1} \in \mathbb{N}_0^m} p(\mathrm{x}_{t-1} | \mathrm{y}_{1:t-1}) p(\mathrm{x}_t | \mathrm{x}_{t-1}) \\
&= \sum_{\mathrm{x}_{t-1} \in \mathbb{N}_0^m} p(\mathrm{x}_{t-1} | \mathrm{y}_{1:t-1}) M_t(\mathrm{x}_{t-1}, \boldsymbol{\eta}(\mathrm{x}_{t-1}), \mathrm{x}_t),
\end{aligned}
$$

Approximating the prediction step

For $\mathrm{x} \in \mathbb{R}^m$ and a length-m probability vector $\boldsymbol{\eta}$, let $M_t(\mathrm{x}, \boldsymbol{\eta}, \cdot)$ be the transition kernel induced by $\mathrm{K}_{t,\boldsymbol{\eta}}$. Then we have:

$$\begin{aligned} p(\mathrm{x}_t | \mathrm{y}_{1:t-1}) &= \sum_{\mathrm{x}_{t-1} \in \mathbb{N}_0^m} p(\mathrm{x}_{t-1} | \mathrm{y}_{1:t-1}) p(\mathrm{x}_t | \mathrm{x}_{t-1}) \\ &\approx \sum_{\mathrm{x}_{t-1} \in \mathbb{N}_0^m} \mathrm{Poi}(\bar{\boldsymbol{\lambda}}_{t-1}) M_t(\mathrm{x}_{t-1}, \boldsymbol{\eta}(\mathbb{E}[\mathrm{x}_{t-1}]), \mathrm{x}_t), \end{aligned}$$

Approximating the prediction step

Lemma

Suppose that $\mathrm{x} \sim \mathrm{Pois}(\boldsymbol{\lambda})$ *for* $\boldsymbol{\lambda} \in \mathbb{R}^m_{\geqslant 0}$ *and* $\bar{x}^{(i)} \sim \mathrm{Bin}(x^{(i)}, \delta^{(i)})$ *for* $\boldsymbol{\delta} \in [0,1]^m$*. Then* $\bar{\mathrm{x}} \sim \mathrm{Pois}(\boldsymbol{\lambda} \odot \boldsymbol{\delta})$*. Furthermore, if* $\mu(\cdot)$ *is the probability mass function associated with* $\mathrm{Pois}(\boldsymbol{\lambda} \odot \boldsymbol{\delta})$ *and* $\mathbb{E}_\mu[\cdot]$ *is the expected value under* μ*, then* $\sum_{\bar{\mathrm{x}} \in \mathbb{N}_0^m} \mu(\bar{\mathrm{x}}) M_t(\bar{\mathrm{x}}, \boldsymbol{\eta}(\mathbb{E}_\mu[\bar{\mathrm{x}}]), \cdot)$ *is the probability mass function associated with* $\mathrm{Pois}\left((\boldsymbol{\lambda} \odot \boldsymbol{\delta})^\top \mathrm{K}_{t,\boldsymbol{\eta}(\boldsymbol{\lambda}\odot\boldsymbol{\delta})}\right)$.

$$p(\mathrm{x}_t | \mathrm{y}_{1:t-1}) \approx \mathrm{Pois}(\boldsymbol{\lambda}_t), \quad \text{with} \quad \boldsymbol{\lambda}_t := (\bar{\boldsymbol{\lambda}}_{t-1} \odot \boldsymbol{\delta}_t)^\top \mathrm{K}_{t,\boldsymbol{\eta}(\bar{\boldsymbol{\lambda}}_{t-1}\odot\boldsymbol{\delta}_t)} + \boldsymbol{\alpha}_t.$$

Approximating the update step

In order to obtain a vector-Poisson approximation to $p(\mathrm{x}_t|\mathrm{y}_{1:t})$ we substitute $\mathrm{Pois}(\boldsymbol{\lambda}_t)$ in place of $p(\mathrm{x}_t|\mathrm{y}_{1:t-1})$ in the update step, which can be viewed as an application of Bayes' rule, and we shall define $\bar{\boldsymbol{\lambda}}_t$ to be the mean vector of the resulting distribution.

Approximating the update step

Lemma

Suppose that $\mathsf{x} \sim \mathrm{Pois}(\boldsymbol{\lambda})$ *for given* $\boldsymbol{\lambda} \in \mathbb{R}^m_{\geqslant 0}$ *and let* $\bar{\mathsf{y}}$ *be a vector with conditionally independent elements distributed* $\bar{y}^{(i)} \sim \mathrm{Bin}(x^{(i)}, q^{(i)})$ *for given* $\mathsf{q} \in [0,1]^m$*. For* G *a row-stochastic* $m \times m$ *matrix and* M *an* $m \times m$ *matrix with rows distributed* $\mathsf{M}^{(i,\cdot)} \sim \mathrm{Mult}(\bar{y}^{(i)}, \mathsf{G}^{(i,\cdot)})$*, let* $\tilde{\mathsf{y}} := \sum_{i=1}^m \mathsf{M}^{(i,\cdot)}$ $\mathsf{y} := \tilde{\mathsf{y}} + \hat{\mathsf{y}}$ *where* $\hat{\mathsf{y}} \sim \mathrm{Pois}(\boldsymbol{\kappa})$ *for a given* $\boldsymbol{\kappa} \in \mathbb{R}^m_{\geqslant 0}$*. Then:*

$$\mathbb{E}[\mathsf{x}|\mathsf{y}] = [1_m - \mathsf{q} + (\{\mathsf{y}^\top \oslash [(\mathsf{q} \odot \boldsymbol{\lambda})^\top \mathsf{G} + \boldsymbol{\kappa}^\top]\}[(1_m \otimes \mathsf{q}) \odot \mathsf{G}^\top])^\top] \odot \boldsymbol{\lambda}.$$

and $\mathsf{y} \sim \mathrm{Pois}([(\boldsymbol{\lambda} \odot \mathsf{q})^\top \mathsf{G}]^\top + \boldsymbol{\kappa})$*.*

Approximating the update step

$$p(\mathsf{x}_t|\mathsf{y}_{1:t}) \approx \mathrm{Poi}(\bar{\boldsymbol{\lambda}}_t)$$

$$\bar{\boldsymbol{\lambda}}_t := [1_m - \mathsf{q}_t + (\{\mathsf{y}_t^\top \oslash [(\mathsf{q}_t \odot \boldsymbol{\lambda}_t)^\top \mathsf{G}_t + \boldsymbol{\kappa}_t^\top]\}[(1_m \otimes \mathsf{q}_t) \odot \mathsf{G}_t^\top])^\top] \odot \boldsymbol{\lambda}_t,$$

Algorithm 1

initialize: $\bar{\boldsymbol{\lambda}}_0 \leftarrow \boldsymbol{\lambda}_0$

for $t \geqslant 1$:

$$\boldsymbol{\lambda}_t \leftarrow [(\bar{\boldsymbol{\lambda}}_{t-1} \odot \boldsymbol{\delta}_t)^\top \mathrm{K}_{t,\boldsymbol{\eta}(\bar{\boldsymbol{\lambda}}_{t-1} \odot \boldsymbol{\delta}_t)}]^\top + \boldsymbol{\alpha}_t$$

$$\bar{\boldsymbol{\lambda}}_t \leftarrow [1_m - \mathrm{q}_t + (\{\mathrm{y}_t^\top \oslash [(\mathrm{q}_t \odot \boldsymbol{\lambda}_t)^\top \mathrm{G}_t + \boldsymbol{\kappa}_t^\top]\}[(1_m \otimes \mathrm{q}_t) \odot \mathrm{G}_t^\top])^\top] \odot \boldsymbol{\lambda}_t$$

$$\boldsymbol{\mu}_t \leftarrow [(\boldsymbol{\lambda}_t \odot \mathrm{q}_t)^\top \mathrm{G}_t]^\top + \boldsymbol{\kappa}_t$$

$$\ell(\mathrm{y}_t | \mathrm{y}_{1:t-1}) \leftarrow -\boldsymbol{\mu}_t^\top 1_m + \mathrm{y}_t^\top \log(\boldsymbol{\mu}_t) - 1_m^\top \log(\mathrm{y}_t!)$$

end for

Consistency

We now allow quantities to depend on a parameter vector $\boldsymbol{\theta} \in \Theta$.
i.e. $\mathsf{K}_{t,\boldsymbol{\eta}}(\boldsymbol{\theta}), \mathsf{q}_t(\boldsymbol{\theta}), \ldots$

Consistency

We now allow quantities to depend on a parameter vector $\boldsymbol{\theta} \in \Theta$.
i.e. $\mathrm{K}_{t,\boldsymbol{\eta}}(\boldsymbol{\theta}), \mathrm{q}_t(\boldsymbol{\theta}), \ldots$

and $\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}), \bar{\boldsymbol{\lambda}}_{t,n}(\boldsymbol{\theta}), \boldsymbol{\mu}_{t,n}(\boldsymbol{\theta})$, and $\ell_n(\boldsymbol{\theta})$

Consistency

We now allow quantities to depend on a parameter vector $\boldsymbol{\theta} \in \Theta$. i.e. $\mathrm{K}_{t,\boldsymbol{\eta}}(\boldsymbol{\theta}), \mathrm{q}_t(\boldsymbol{\theta}), \ldots$

and $\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}), \bar{\boldsymbol{\lambda}}_{t,n}(\boldsymbol{\theta}), \boldsymbol{\mu}_{t,n}(\boldsymbol{\theta})$, and $\ell_n(\boldsymbol{\theta})$

Asymptotic analysis is under the large population regime $n \to \infty$ and we fix a time horizon $T < \infty$.

Consistency

We now allow quantities to depend on a parameter vector $\boldsymbol{\theta} \in \Theta$. i.e. $\mathrm{K}_{t,\boldsymbol{\eta}}(\boldsymbol{\theta}), \mathrm{q}_t(\boldsymbol{\theta}), \ldots$

and $\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}), \bar{\boldsymbol{\lambda}}_{t,n}(\boldsymbol{\theta}), \boldsymbol{\mu}_{t,n}(\boldsymbol{\theta})$, and $\ell_n(\boldsymbol{\theta})$

Asymptotic analysis is under the large population regime $n \to \infty$ and we fix a time horizon $T < \infty$.

Fix a 'ground truth' parameter $\boldsymbol{\theta}^* \in \Theta$

Law of Large Numbers

we show that for certain deterministic vectors $\boldsymbol{\nu}_t(\boldsymbol{\theta}^*)$ and $\boldsymbol{\mu}_t(\boldsymbol{\theta}^*)$, $t \geqslant 1$,

$$\frac{1}{n}\mathrm{x}_t \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\nu}_t(\boldsymbol{\theta}^*),$$
$$\frac{1}{n}\mathrm{y}_t \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\mu}_t(\boldsymbol{\theta}^*).$$

We find deterministic vectors $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ and $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, $t \geqslant 1$, $\boldsymbol{\theta} \in \Theta$, where $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ is a function of $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, such that:

$$\frac{1}{n}\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\theta^*} \boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}), \qquad \frac{1}{n}\boldsymbol{\mu}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\theta^*} \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}).$$

We find deterministic vectors $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ and $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, $t \geqslant 1$, $\boldsymbol{\theta} \in \Theta$, where $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ is a function of $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, such that:

$$\frac{1}{n}\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}), \qquad \frac{1}{n}\boldsymbol{\mu}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}).$$

It turns out that:

$$\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \boldsymbol{\nu}_t(\boldsymbol{\theta}^*),$$
$$\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \boldsymbol{\mu}_t(\boldsymbol{\theta}^*).$$

We find deterministic vectors $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ and $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, $t \geqslant 1$, $\boldsymbol{\theta} \in \Theta$, where $\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ is a function of $\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})$, such that:

$$\frac{1}{n}\boldsymbol{\lambda}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}), \qquad \frac{1}{n}\boldsymbol{\mu}_{t,n}(\boldsymbol{\theta}) \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}).$$

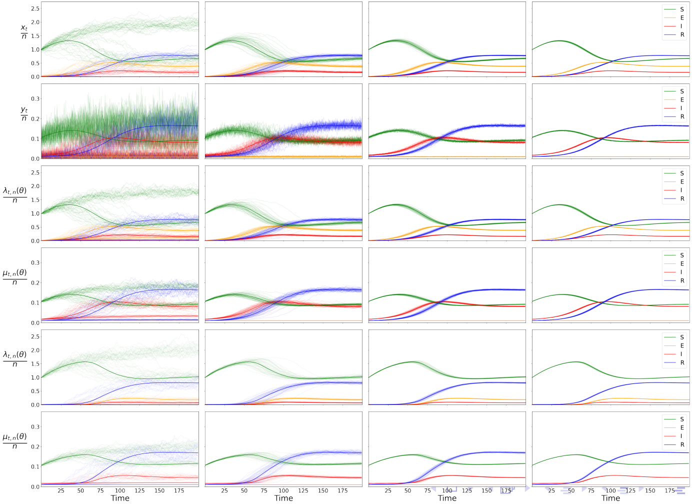
It turns out that:

$$\boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \boldsymbol{\nu}_t(\boldsymbol{\theta}^*),$$
$$\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) = \boldsymbol{\mu}_t(\boldsymbol{\theta}^*).$$

So that, in this sense, we achieve asymptotically exact filtering:

$$n^{-1}\mathrm{x}_t \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\lambda}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*),$$
$$n^{-1}\mathrm{y}_t \xrightarrow[a.s.]{\boldsymbol{\theta}^*} \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*).$$

Theory illustration



Consistency of the maximum PAL estimator

$$\frac{1}{n}\ell_n(\boldsymbol{\theta}) - \frac{1}{n}\ell_n(\boldsymbol{\theta}^*) \xrightarrow[a.s.]{\boldsymbol{\theta}^*} -\sum_{t=1}^{T} \mathrm{KL}\left(\mathrm{Poi}[\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*)] \,\|\, \mathrm{Poi}[\boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta})]\right).$$

Theory illustration

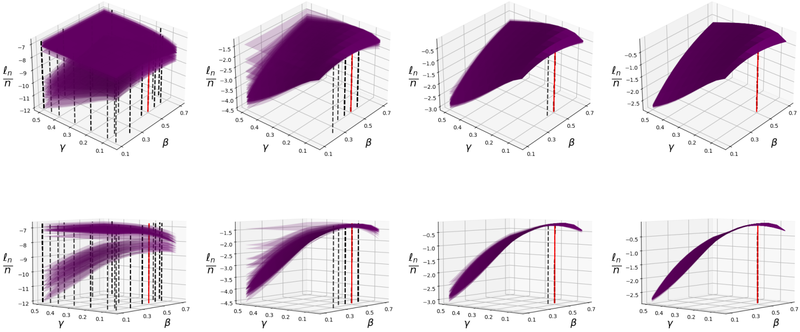


Figure: Simulation SEIR example.

Consistency

Theorem

Let some assumptions hold and let $\hat{\boldsymbol{\theta}}_n$ be a maximiser of $\ell_n(\boldsymbol{\theta})$. Then $\hat{\boldsymbol{\theta}}_n$ converges to Θ^* as $n \to \infty$, $\mathbb{P}^{\boldsymbol{\theta}^*}$-almost surely.

(Whitehouse et al., 2022)

Consistency

Theorem

Let some assumptions hold and let $\hat{\boldsymbol{\theta}}_n$ be a maximiser of $\ell_n(\boldsymbol{\theta})$. Then $\hat{\boldsymbol{\theta}}_n$ converges to Θ^* as $n \to \infty$, $\mathbb{P}^{\boldsymbol{\theta}^*}$-almost surely.

(Whitehouse et al., 2022)

$$\Theta^* := \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) \text{ for all } t = 1, \ldots T\}$$

Consistency

Theorem

Let some assumptions hold and let $\hat{\boldsymbol{\theta}}_n$ be a maximiser of $\ell_n(\boldsymbol{\theta})$. Then $\hat{\boldsymbol{\theta}}_n$ converges to Θ^* as $n \to \infty$, $\mathbb{P}^{\boldsymbol{\theta}^*}$-almost surely.

(Whitehouse et al., 2022)

$$\Theta^* := \{\boldsymbol{\theta} \in \Theta : \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}) = \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) \text{ for all } t = 1, \ldots T\}$$

Identifiability:

Lemma

$$\boldsymbol{\theta} \in \Theta^* \iff \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}, \boldsymbol{\theta}) = \boldsymbol{\mu}_{t,\infty}(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*), \quad \forall t = 1, \ldots, T,$$

PAL applications

- Can be embedded within a delayed acceptance particle mcmc scheme to speed up exact bayesian inference
- Can be used within Stan
- Used to fit a large scale meta population model of measles

Whitehouse et al. (2022)

Whitehouse, M., Whiteley, N., and Rimella, L. (2022). Consistent and fast inference in compartmental models of epidemics using poisson approximate likelihoods. *arXiv preprint arXiv:2205.13602*.

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.

Thanks, any Questions?