Pandemic Intensity Estimation from Stochastic Approximation-based Algorithms

P. Abry[1], J. Chevallier[2], G. Fort[2], B. Pascal[3]

[1] LPENSL & CNRS, Lyon, France [2]IMT & INSA and CNRS, Toulouse, France [3] LS2N & CNRS, Nantes, France

Work partially supported by the *Fondation Simone et Cino Del Duca* and the *French National Research Agency.*

Aims and Methods

Intensity monitoring of a pandemic is critical to design sanitary policies and to quantify their effectiveness. The task is significantly complicated by the low quality of reported infection counts and by the need for daily updates while the pandemic is still active.

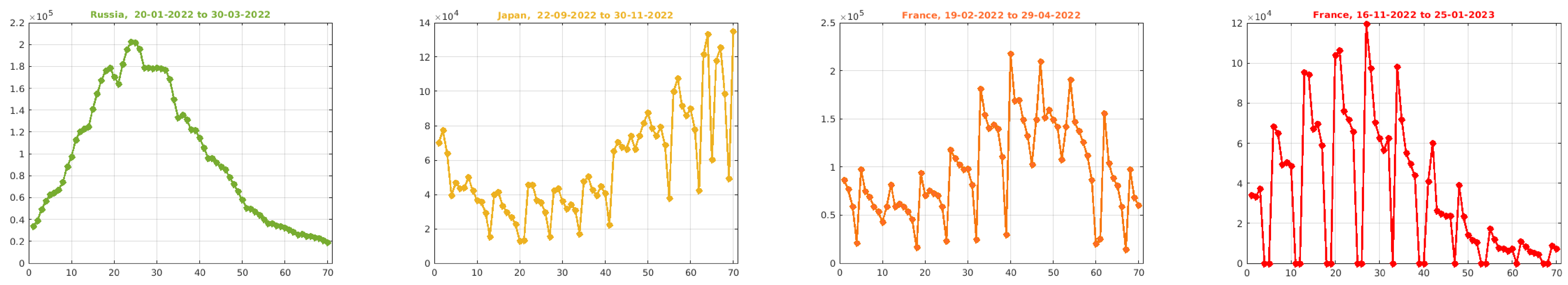
Strategies were devised, focusing on a **time-varying reproduction number** R_t: defined as *the average number of secondary cases of disease caused by a single infected individual over his/her infectious period.* It characterizes pathogen transmissibility during an epidemic (Cori et al, 2013).

Contributions:

- A parametric Hidden Markov model.
- Data-driven design of the parameters, based on Monte Carlo methods and Stochastic Optimization.
- Point estimate and credibility interval estimate of R_t.

Covid19 data (from Johns Hopkins University)

Daily counts $(\mathsf{Z}_t)_t$ of new infections, 70 successive days



An inverse problem (Pascal et al. 2022) adapted from (Cori et al. 2013)

Data fidelity term: given the distribution $(\phi_u)_u$ of the *serial interval* i.e. time between the onset of symptoms in a primary case and the onset of symptoms in secondary cases,

$$\mathsf{Z}_t|\text{past}_{t-1} \sim \mathcal{P}\left(\mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t)\right) \qquad \mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t) \stackrel{\text{def}}{=} \mathsf{R}_t \sum_{u=1}^{\tau} \phi_u \mathsf{Z}_{t-u} + \mathsf{O}_t \qquad \phi_u \equiv \text{pdf Gamma}\left(\begin{smallmatrix}\text{mean } 6.68\\ \text{std } 3.53\end{smallmatrix}\right)$$

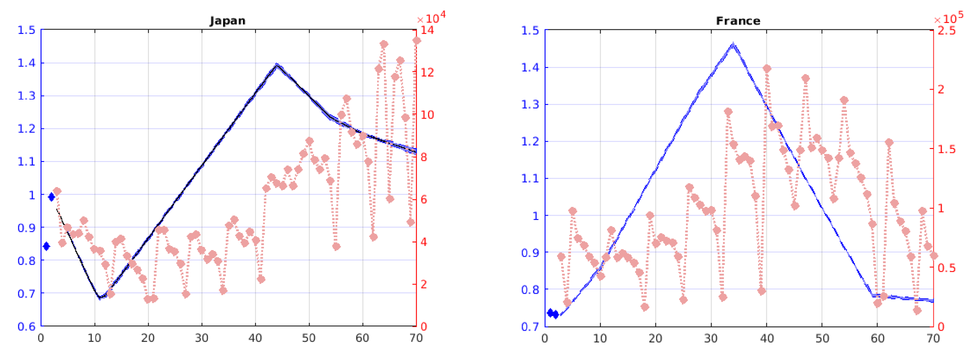
Penalty term: sparse second derivative of the time evolution of R_t, sparse errors O_t

Point estimate: under constraints $\mathsf{R}_t \geq 0$ and $\mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t) \geq 0$, solve

$$\text{argmin}_{\mathsf{R}_{1:T}, \mathsf{O}_{1:T}} \sum_{t=1}^{T} \left(\mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t) - \mathsf{Z}_t \ln \mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t)\right) + \lambda_{\mathsf{R}} \|\mathsf{D}\, \mathsf{R}_{1:T}\|_1 + \lambda_{\mathsf{O}} \|\mathsf{O}_{1:T}\|_1$$

Estimates of R_t for fixed $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$

95% credibility intervals and posterior mean



$(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) = (1.05e5, 0.05)$ $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}) = (1.48e5, 0.05)$

A Hidden Markov model (Fort et al, 2023)

Hidden processes: $(\mathsf{R}_t)_t$ and $(\mathsf{O}_t)_t$ are independent

$$\mathsf{R}_t|\mathsf{R}_{-1:t-1} \sim 2\mathsf{R}_{t-1} - \mathsf{R}_{t-2} + \text{Laplace}\left(\lambda_{\mathsf{R}}/4\right) \qquad \mathsf{O}_t|\mathsf{O}_{-1:t-1} \sim \text{Laplace}\left(\lambda_{\mathsf{O}}\right)$$

Observation process: $\mathsf{Z}_t|\text{past}_{t-1} \sim \mathcal{P}\left(\mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t)\right)$ when $\mathsf{R}_t \geq 0$ and $\mathcal{I}_t(\mathsf{R}_t, \mathsf{O}_t) \geq 0$

Point estimates and Credibility interval estimates of R_t**:** from a Markov chain Monte Carlo approximation of the *posterior probability distribution* $\pi\left(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}|\mathsf{Z}_{1:T}; \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}\right)$

Estimates of R_t when $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$ maximize the likelihood

Maximum Likelihood criterion

$$\text{argmax}_{\lambda_{\mathsf{R}}>0, \lambda_{\mathsf{O}}>0} \int \pi\left(\mathsf{Z}_{1:T}, \mathsf{R}_{1:T}, \mathsf{O}_{1:t}; \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}}\right)\, \mathsf{dR}_{1:T}\, \mathsf{dO}_{1:T}$$

Expectation Maximization: (Dempster et al, 1977)

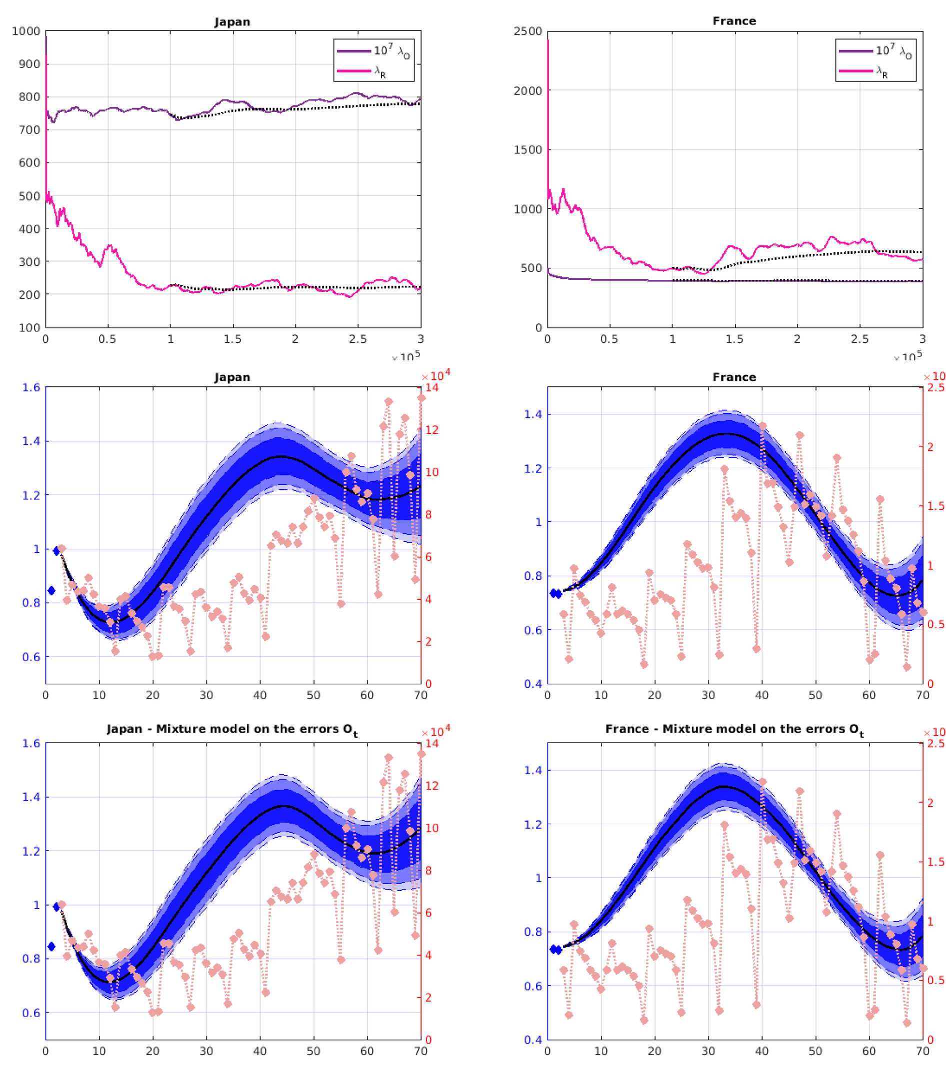
- the E-step

$$\mathsf{S}_{\mathsf{R}}(\lambda_\bullet^{(k)}) \stackrel{\text{def}}{=} \int \|\mathsf{D}\,\mathsf{R}_{1:T}\|_1\, \pi\left(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}|\mathsf{Z}_{1:T}; \lambda_\bullet^{(k)}\right)\, \mathsf{dR}_{1:T}\, \mathsf{dO}_{1:T}$$

$$\mathsf{S}_{\mathsf{O}}(\lambda_\bullet^{(k)}) \stackrel{\text{def}}{=} \int \|\mathsf{O}_{1:T}\|_1\, \pi\left(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}|\mathsf{Z}_{1:T}; \lambda_\bullet^{(k)}\right)\, \mathsf{dR}_{1:T}\, \mathsf{dO}_{1:T}$$

- the M-step

$$\lambda_{\mathsf{R}}^{(k+1)} \stackrel{\text{def}}{=} \frac{T}{\mathsf{S}_{\mathsf{R}}(\lambda_\bullet^{(k)})} \qquad \lambda_{\mathsf{O}}^{(k+1)} \stackrel{\text{def}}{=} \frac{T}{\mathsf{S}_{\mathsf{O}}(\lambda_\bullet^{(k)})}.$$



Stochastic Approximation EM: (Delyon et al, 1999)

- the E-step

$$S_{\mathsf{R}}^{(k+1)} \stackrel{\text{def}}{=} S_{\mathsf{R}}^{(k)} + \gamma_{\mathsf{R},k+1}\left(\widehat{\mathsf{S}_{\mathsf{R}}}^{(k+1)} - S_{\mathsf{R}}^{(k)}\right)$$

$$S_{\mathsf{O}}^{(k+1)} \stackrel{\text{def}}{=} S_{\mathsf{O}}^{(k)} + \gamma_{\mathsf{O},k+1}\left(\widehat{\mathsf{S}_{\mathsf{O}}}^{(k+1)} - S_{\mathsf{O}}^{(k)}\right)$$

$\widehat{\mathsf{S}_{\mathsf{U}}}^{(k+1)}$ is a Monte Carlo approximation of $\mathsf{S}_{\mathsf{U}}(\lambda_\bullet^{(k)})$.

- the M-step

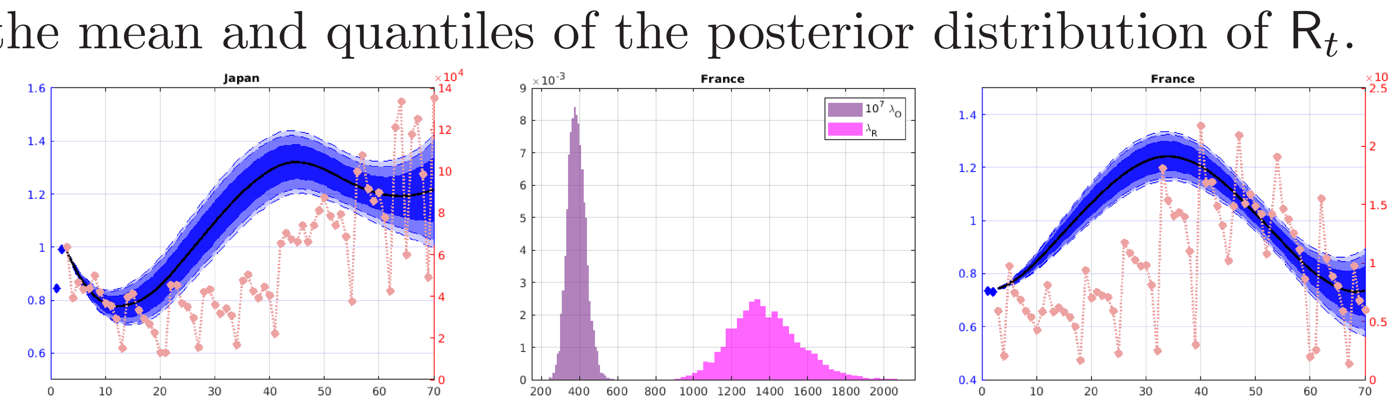
$$\lambda_{\mathsf{R}}^{(k+1)} \stackrel{\text{def}}{=} \frac{T}{S_{\mathsf{R}}^{(k+1)}} \qquad \lambda_{\mathsf{O}}^{(k+1)} \stackrel{\text{def}}{=} \frac{T}{S_{\mathsf{O}}^{(k+1)}}.$$

Estimates of R_t by integration w.r.t. $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$

Prior on $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$**:** a flat prior.

Joint posterior distribution of $(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}, \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$**:** proportional to $\pi(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}|\mathsf{Z}_{1:T}; \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$.

Point estimates and Credibility interval estimates of R_t**:** from a Markov chain Monte Carlo approximation of the *joint posterior distibution* of $(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}, \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$ given $\mathsf{Z}_{1:T}$, compute the mean and quantiles of the posterior distribution of R_t.



Conclusion

- Efficient MCMC methods for non-smooth concave density with a support: when optimization and Langevin-based Monte Carlo sampling intertwine
- Sequential Monte Carlo methods for filtering, smoothing and prediction
- Validation of the credibility intervals
- Stable hidden process

References

- Cori et al, *A new framework anf software to estimate time-varying reproduction numbers during epidemics*, Am. J. Epidemiol., 2013.
- Delyon et al, *Convergence of a Stochastic Approximation version of the EM algorithm*, Ann. Stat., 1999.
- Dempster et al, *Maximum Likelihood from Incomplete Data via the EM algorithm*, J. Roy. Stat. Soc. B Meth., 1977.
- Fort et al, *Covid19 Reproduction number: Credibility Intervals by Blockwise Proximal Monte Carlo samplers*, IEEE Trans. Signal Process., 2023.
- Pascal et al, *Nonsmooth convex optimization to estimate the Covid19 reproduction number space-time evolution with robustness against low quality data*, IEEE Trans. Signal Process., 2022.