Credibility intervals for pandemic reproduction number from Nonsmooth Langevin-type Monte Carlo sampling

Juliette Chevallier

INSA Toulouse
Institut de Mathématiques de Toulouse

Gersende Fort

CNRS
IMT & Laboratoire d'Analyse et d'Architecture des Systèmes

Séminaire Maths-Bio-Santé. December 2024

In collaboration with

- Patrice Abry, CNRS, Lab. de Physique de l'ENS Lyon, France
- Barbara Pascal, CNRS, LS2N, Nantes, France
- Nelly Pustelnik, CNRS, Lab. de Physique de l'ENS Lyon, France

Talk based on the papers:

- *Sampling Nonsmooth Log-Concave Densities: A Comparative Study of Primal-Dual Based Proposal Distributions*. by J. Chevallier and G. Fort Preprint 2024 (HAL 04824190)
- *Hierarchical Bayesian Estimation of COVID-19 Reproduction Number* by P. Abry, J. Chevallier, G. Fort and B. Pascal. Preprint 2024 (HAL 04695138)
- *Pandemic Intensity Estimation from Stochastic Approximation-Based Algorithms* by P. Abry, J. Chevallier, G. Fort and B. Pascal. CAMSAP 2023 (HAL 04174245)
- *Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo samplers* by G. Fort, B. Pascal, P. Abry and N. Pustelnik. IEEE Trans. Signal Proc. 2023 (HAL 03611079)
- *Temporal evolution of the Covid19 pandemic reproduction number: Estimations from Proximal optimization to Monte Carlo sampling* by P. Abry, G. Fort, B. Pascal and N. Pustelnik. EMBC 2022 (HAL 03565440)
- *Credibility intervals design for Covid19 reproduction number from nonsmooth Langevin-type Monte Carlo sampling* by H. Artigas, B. Pascal, G. Fort, P. Abry and N. Pustelnik. EUSIPCO 2022 (HAL 03371837)
- *Estimation et intervalles de crédibilité pour le taux de reproduction de la Covid19 par échantillonnage Monte Carlo Langevin proximal* by P. Abry, G. Fort, B. Pascal and N. Pustelnik. GRETSI 2022 (HAL 03611891)

Partly funded by Fondation Del Duca (OpSiMorE) and **by** ANR (OptiMoCSI)

Outline of the talk

- Part I: Reproduction number of the Covid19

1. The problem at hand, the Data
2. The Model
3. Estimation of the R_t's
4. Conclusions: works in progress, the future

- Part II: Sampling a non smooth log-concave density

I. Reproduction number of the Covid19
The Data

Credibility intervals for the Reproduction number R, why ?

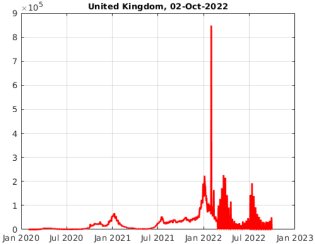
- Monitoring the Covid19 pandemic constitutes a critical societal stake: Covid19 pandemic caused/is causing unprecedented health, social, and economic crises.
- Need to assess the intensity of the/a pandemic, prerequisite for efficient sanitary policies.

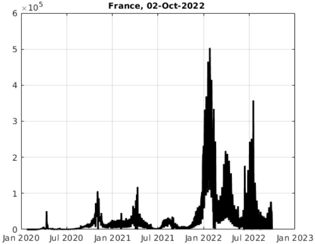
- The reproduction number measures
  - the strength of the pandemic by quantifying rate of growth of daily new infections
  - the number of second infections caused by one primary infection.

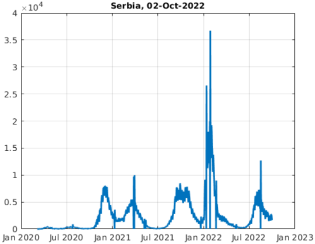
- Estimation of the daily R_t
  - by a value of the index
  - by credibility intervals: valuable information for the decision makers, notably in periods of rapid evolution or of changes in trends.

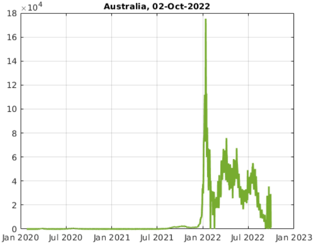
The data: daily new infections

- Real data, from Johns Hopkins University repository
- Examples for UK, France, Serbia and Australia



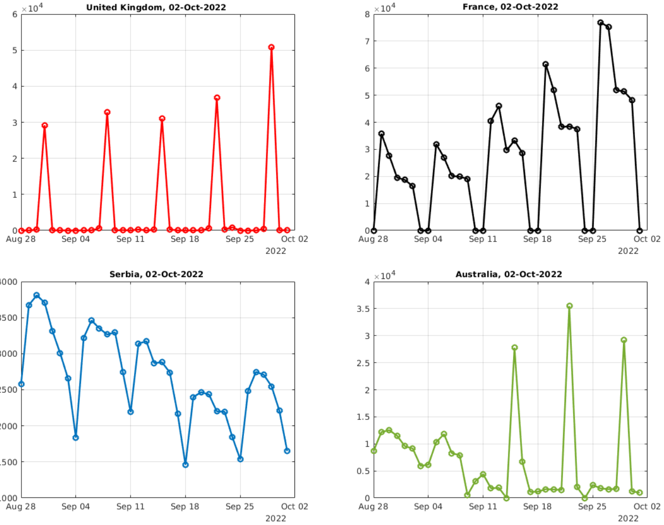






The data: daily new infections - zoom on a 35-day period of 2022

- Examples for UK, France, Serbia and Australia



I. Reproduction number of the Covid19
The Model

Epidemiological model

From Cori et al Cori et al (2013)

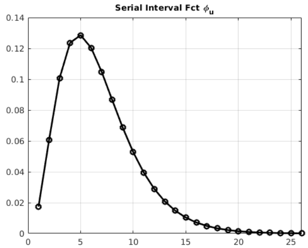
- The data $Z_1, \cdots, Z_T$: non negative integers
- Parameter: $(\mathsf{R}_1, \cdots, \mathsf{R}_T) \in (\mathbb{R}_+)^T$
- Conditionally to the past

$$Z_t \,\Big|\, Z_{1:(t-1)} \sim \mathcal{P}\left(\mathsf{R}_t\, \Phi_t^{\mathsf{Z}}\right) \qquad \text{where } \Phi_t^{\mathsf{Z}} := \sum_{u=1}^{\tau_\phi} \phi_u Z_{t-u}$$

- $\tau_\phi = 26$ days
- $\phi_u := \mathrm{PDF}_{\mathrm{Gamma}}(u)$

shape $= 1/0.28$, scale $= 1.87$

mean: 6.68 days
std: 3.53 days
mode: 4.8 days



Epidemiological model $\rightsquigarrow$ Maximum Likelihood Estimator (MLE)

From Cori et al Cori et al (2013)

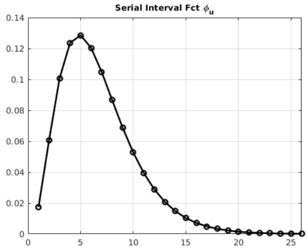
- The data $Z_1, \cdots, Z_T$: non negative integers
- *Parameter*: $(\mathsf{R}_1, \cdots, \mathsf{R}_T) \in (\mathbb{R}_+)^T \quad \rightsquigarrow \quad \widehat{\mathsf{R}}_t := \frac{Z_t}{\Phi_t^{\mathsf{Z}}}$
- Conditionally to the past

$$Z_t \,\Big|\, Z_{1:(t-1)} \sim \mathcal{P}\left(\mathsf{R}_t\, \Phi_t^{\mathsf{Z}}\right) \qquad \text{where } \Phi_t^{\mathsf{Z}} := \sum_{u=1}^{\tau_\phi} \phi_u Z_{t-u}$$

- $\tau_\phi = 26$ days
- $\phi_u := \mathrm{PDF}_{\mathrm{Gamma}}(u)$

shape $= 1/0.28$, scale $= 1.87$

mean: 6.68 days
std: 3.53 days
mode: 4.8 days

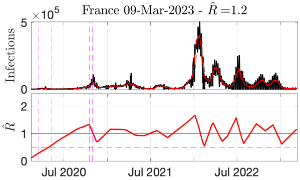


Bayesian framework

We move to a Bayesian setting and consider the R_t's as latent variables

- A priori distribution on the R_t's provides regularization
  Required here: as many parameters as observations
- Credibility interval through quantiles of the marginal distributions of the R_t's
- Maximum A Posteriori (MAP) Estimation of the R_t's

First Bayesian model



From Abry et al Abry et al (2020)

- A priori distribution on the R_t's

to ensure piecewise linear time evolutions of $t \mapsto \mathsf{R}_t$

- L^1 penalization of the discrete time second derivative of $t \mapsto \mathsf{R}_t$

$$\ln \text{prior} := -\lambda_{\mathsf{R}} \, \|\mathsf{D}\, \mathsf{R}_{1:T} + \delta\|_1 \qquad \text{up to an additive constant}$$

where

$$\mathsf{D} := \frac{1}{4}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ -2 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & \cdots & & 0 & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{T\times T} \qquad \delta := \frac{1}{4}\begin{bmatrix} \mathsf{R}_{-1} - 2\mathsf{R}_0 \\ \mathsf{R}_0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \in \mathbb{R}^T$$

- $\lambda_{\mathsf{R}} := 3.5 \times \mathrm{std}(Z_1, \cdots, Z_T)$

$$\mathsf{R}_{1:T} := \begin{bmatrix} \mathsf{R}_1 \\ \mathsf{R}_2 \\ \cdots \\ \mathsf{R}_T \end{bmatrix} \in \mathbb{R}^T$$

Not robust against low-quality data

Robust Bayesian model

From Pascal et al Pascal et al (2021)

- Consider Cori et al. model + regularized R_t's i.e. previous model

- Model the errors on the counts via $\mathsf{O}_1, \cdots, \mathsf{O}_T$ in $\mathbb{R}^T$

- corrupted data, with pseudo-seasonalities, under-evaluations/over-evaluations ($\rightarrow$ negative counts)
- a priori distribution **and** modification of the likelihood

- Given initial value $\mathcal{I} := (\mathsf{R}_{-1}, \mathsf{R}_0, Z_{-\tau_\phi+1}, \ldots, Z_0)$

$$Z_t \Big| \mathsf{R}_t, \mathsf{O}_t, Z_{1:(t-1)}, \mathcal{I} \sim \mathcal{P}\left(\mathsf{R}_t \, \Phi_t^{\mathsf{Z}} + \mathsf{O}_t\right) \qquad \text{where } \Phi_t^{\mathsf{Z}} := \sum_{u=1}^{\tau_\phi} \phi_u Z_{t-u}$$

$$\ln \text{prior} := -\lambda_{\mathsf{R}} \, \|\mathsf{D}\, \mathsf{R}_{1:T} + \delta\|_1 - \lambda_{\mathsf{O}} \, \|\mathsf{O}_{1:T}\|_1 \qquad \text{up to an additive constant}$$

- A constraint set

$$\mathcal{D} := \bigcap_t \left\{ (\mathsf{R}_t, \mathsf{O}_t) \text{ s.t } \mathsf{R}_t \geq 0 \text{ and } \mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t \geq 0 \right\}$$

- $\lambda_{\mathsf{O}} := 0.05$

Parametric Hidden Markov model (HMM)

From Fort et al Fort et al (2023)

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

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

- **Observation process**: Given past and initial value $\mathcal{I}$

$$Z_t \,\Big|\, \mathsf{R}_t, \mathsf{O}_t, Z_{1:(t-1)}, \mathcal{I} \;\sim\; \mathcal{P}\left(\mathsf{R}_t\,\Phi_t^{\mathsf{Z}} + \mathsf{O}_t\right) \quad \text{when } \mathsf{R}_t \geq 0 \text{ and } \mathsf{R}_t\Phi_t^{\mathsf{Z}} + \mathsf{O}_t \geq 0$$

- **Point estimates** and **Credibility interval** estimates of R_t from the a posteriori distribution of

$$\theta := (\mathsf{R}_{1:T}, \mathsf{O}_{1:T}) \in (\mathbb{R}_+)^T \times \mathbb{R}^T$$

given $Z_1, \cdots, Z_T$, initial value $\mathcal{I}$ and $(\lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$

I. Reproduction number of the Covid19
Estimation of the R_t's

Bayesian estimation: Point estimates

- Quantiles and other statistics: **Maximum a Posteriori**, mean a posteriori, etc

- for each component R_t and O_t
- based on the marginal distributions of the distribution of $\theta := (\mathsf{R}_{1:T}, \mathsf{O}_{1:T})$

- From the joint distribution $\pi(Z_{1:T}, \theta | \lambda_\mathsf{R}, \lambda_\mathsf{O})$, obtain the a posteriori distribution $\pi(\theta | Z_{1:T}, \lambda_\mathsf{R}, \lambda_\mathsf{O})$ of the form

$$- \ln \text{posterior:} \quad \theta \mapsto \begin{cases} f(\theta) + h(\mathsf{A}\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D} & 0_{T\times T} \\ 0_{T\times T} & \mathrm{I}_T \end{bmatrix} \in \mathbb{R}^{2T\times 2T}$$

Given $Z_{1:T}$, $\mathcal{I}$, λ_O and λ_R, up to an additive constant:

$$\begin{cases} f(\theta) := \sum_{t=1}^{T} \left\{ (\mathsf{R}_t \Phi_t^\mathsf{Z} + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t^\mathsf{Z} + \mathsf{O}_t) \right\} \\ h(\mathsf{A}\theta) := \lambda_\mathsf{R} \, \|\mathsf{D}\,\mathsf{R}_{1:T} + \delta\|_1 + \lambda_\mathsf{O} \, \|\mathsf{O}_{1:T}\|_1 - T \ln \lambda_\mathsf{R} - T \ln \lambda_\mathsf{O} \end{cases}$$

- Here: $\lambda_\mathsf{R} = 3.5 \times \mathsf{std}(Z_{1:T})$ and $\lambda_\mathsf{O} = 0.05$

Maximum a Posteriori estimation of the (R_t, O_t)'s

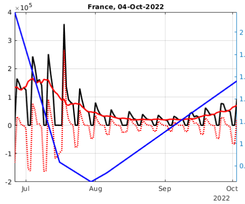
- Does MAP exist ? Unique ? Pascal et al (2022), Fort et al (2023)
  - If $\Phi_t^Z > 0$ and $\Phi_{t'}^Z > 0$ for $t < t' \leq T$, and $Z_{t''} > 0$ then a MAP exists.
  - It two MAP, then: same Poisson intensity $R_t \Phi_t^Z + O_t = R_{t'} \Phi_{t'}^Z + O_{t'}$ and same sign $O_t O_t' \geq 0; (DR_t)\,(DR_t') \geq 0$.

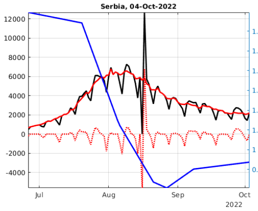
- Computation: a **Chambolle-Pock** iterative algorithm proposed by Pascal et al (2022); see also Abry et al (2020)

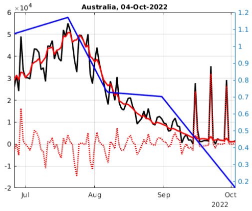
- MAP for France, Serbia and Australia over the last 100 days:

- `left y-axis`: the data Z_t, (dots) $\hat{O}_t$ by MAP and (line) $Z_t - \hat{O}_t$

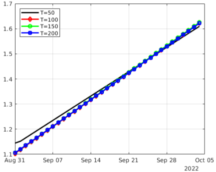
- `right y-axis`: $\hat{R}_t$ by MAP.







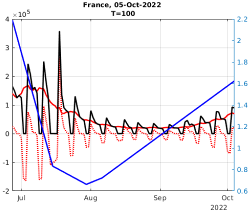
MAP estimation of the (R_t, O_t)'s: Influence of the parameters

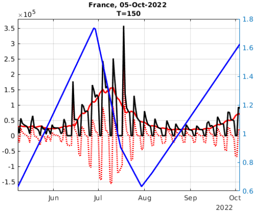


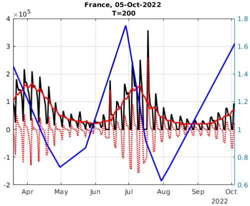
- Role of T for the MAP estimate of the last R_t's:

- `below`: MAP estimate computed for $T = 100$, $T = 150$ and $T = 200$ observations,

- `right`: The three estimates for the last 35 days.





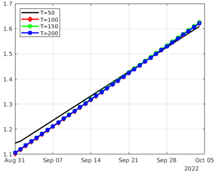


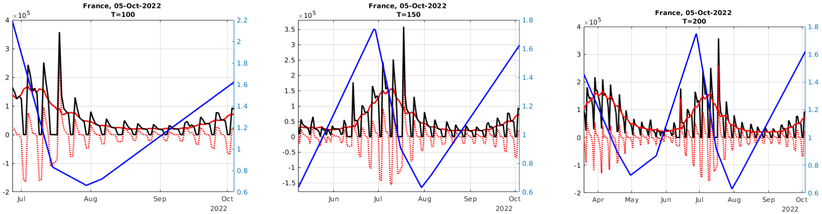
MAP estimation of the $(\mathrm{R}_t, \mathrm{O}_t)$'s: Influence of the parameters

- Role of T for the MAP estimate of the last R_t's:

- `below`: MAP estimate computed for $T = 100$, $T = 150$ and $T = 200$ observations,

- `right`: The three estimates for the last 35 days.





- Role of λ_{R} and λ_{O} for the MAP estimate:
  - λ_{O} and $\lambda_{\mathrm{R}}/\Phi_t^{\mathrm{Z}}$ quantify how much the MAP is allowed to differ from the MLE
  - Small $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$ implie Laplace a priori for O_t and $[\mathrm{DR}]_t$ with large variance

$\rightsquigarrow$ *Importance of the calibration of λ_{O} and λ_{R}*

Two contributions

- Maximum likelihood criterion: From Abry et al (2023)

$$\underset{\lambda_{\mathsf{R}}>0,\lambda_{\mathsf{O}}>0}{\operatorname{argmax}} \; \pi\left(Z_{1:T}|\lambda_{\mathsf{R}},\lambda_{\mathsf{O}}\right) \;=\; \underset{\lambda_{\mathsf{R}}>0,\lambda_{\mathsf{O}}>0}{\operatorname{argmax}} \int_{\mathcal{D}} \pi\left(Z_{1:T},\theta|\lambda_{\mathsf{R}},\lambda_{\mathsf{O}}\right) \,\mathsf{d}\theta$$

and then, consider $\theta \mapsto \pi(\theta|Z_{1:T},\lambda_{\mathsf{R}}^{\mathsf{MLE}},\lambda_{\mathsf{O}}^{\mathsf{MLE}}) \propto \pi(Z_{1:T},\theta|\lambda_{\mathsf{R}}^{\mathsf{MLE}},\lambda_{\mathsf{O}}^{\mathsf{MLE}})$

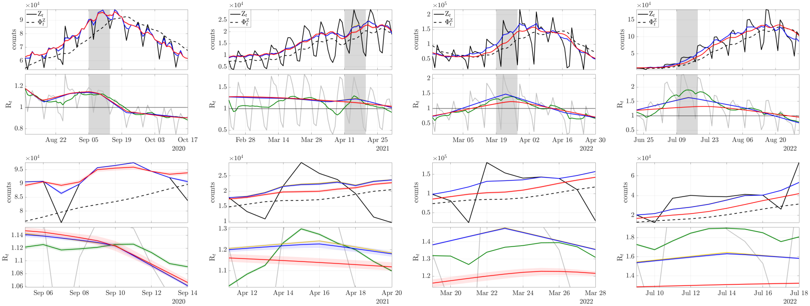
- A Full Bayesian approach: cancel the dependence upon $(\lambda_{\mathsf{R}},\lambda_{\mathsf{O}})$ by weighted integration From Abry et al. (2024)

$$\pi^\star(Z_{1:T},\theta) := \int_{(\mathbb{R}_+)^2} \pi(Z_{1:T},\theta|\lambda_{\mathsf{R}},\lambda_{\mathsf{O}})\,\mathsf{prior}(\lambda_{\mathsf{R}},\lambda_{\mathsf{O}})\;\mathsf{d}\lambda_{\mathsf{R}}\,\mathsf{d}\lambda_{\mathsf{O}}$$

and then, consider $\theta \mapsto \pi^\star(\theta|Z_{1:T}) \propto \pi^\star(Z_{1:T},\theta)$

Role of $(\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}})$ on point estimates of the $(\mathrm{R}_t, \mathrm{O}_t)$'s

From Abry et al (2024)



(a) India (b) Germany (c) France (d) South Korea

$\hat{\theta}$ via MLE (no $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$); $\hat{\theta}$ via "state of the art" (no $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$);
$\hat{\theta}$ via MAP (fixed $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$); $\hat{\theta}$ via mean a posteriori (fixed $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$);
$\hat{\theta}$ via full Bayesian and mean a posteriori (integrated w.r.t. $\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}$)

First and third rows: COVID-19 daily new infection counts and infectiouness (solid and dashed black curves respectively).
Second and fourth rows: estimated reproduction number.

Top rows: $T = 70$-day whole period. Bottom rows: zoom on the ten-day period shaded in gray on top rows.

Bayesian estimation: Credibility intervals

- A posteriori distribution of the form

$$-\ln \text{posterior:} \quad \theta \mapsto \begin{cases} f(\theta) + h(\mathsf{A}\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D} & 0_{T\times T} \\ 0_{T\times T} & \mathrm{I}_T \end{bmatrix} \in \mathbb{R}^{2T\times 2T}$$

Given $Z_{1:T}$, $\mathcal{I}$, λ_{O} and λ_{R}, up to an additive constant:

$$\begin{cases} f(\theta) := \sum_{t=1}^{T} \left\{ (\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) \right\} \\ h(\mathsf{A}\theta) := \lambda_{\mathsf{R}} \, \|\mathsf{D}\, \mathsf{R}_{1:T} + \delta\|_1 + \lambda_{\mathsf{O}} \, \|\mathsf{O}_{1:T}\|_1 - T \ln \lambda_{\mathsf{R}} - T \ln \lambda_{\mathsf{O}} \end{cases}$$

- Credibility intervals through quantiles $\longrightarrow$ empirical quantiles

$\rightsquigarrow$ We need to know how to sample R_t's and O_t's according to this distribution

Bayesian estimation: Credibility intervals via MCMC

- A posteriori distribution of the form

$$-\ln \text{posterior}: \quad \theta \mapsto \begin{cases} f(\theta) + h(\mathsf{A}\,\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases} \qquad \mathsf{A} := \begin{bmatrix} \mathsf{D} & 0_{T\times T} \\ 0_{T\times T} & \mathrm{I}_T \end{bmatrix} \in \mathbb{R}^{2T\times 2T}$$

Given $Z_{1:T}$, $\mathcal{I}$, λ_{O} and λ_{R}, up to an additive constant:

$$\begin{cases} f(\theta) := \sum_{t=1}^{T} \left\{ (\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t^{\mathsf{Z}} + \mathsf{O}_t) \right\} \\ h(\mathsf{A}\theta) := \lambda_{\mathsf{R}} \, \|\mathsf{D}\,\mathsf{R}_{1:T} + \delta\|_1 + \lambda_{\mathsf{O}} \, \|\mathsf{O}_{1:T}\|_1 - T \ln \lambda_{\mathsf{R}} - T \ln \lambda_{\mathsf{O}} \end{cases}$$

- Credibility intervals through quantiles $\longrightarrow$ empirical quantiles

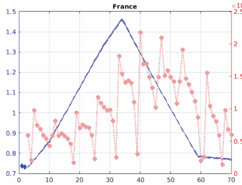
$\rightsquigarrow$ We need to know how to sample R_t's and O_t's according to this distribution

Markov Chain Monte Carlo samplers

Assume such a sampler is well defined $\rightsquigarrow$ Cf. part II

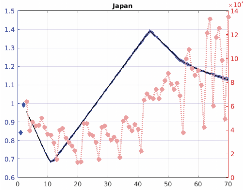
Summary: 95% credibility intervals and posterior mean

Fixed values

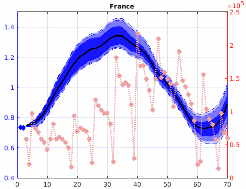


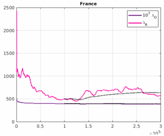
$$\begin{cases} \lambda_R^{\text{France}} = 1.05 \times 10^5 \\ \lambda_O = 0.05 \end{cases}$$

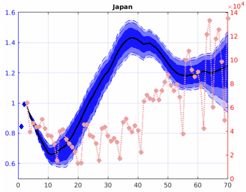
$$\begin{cases} \lambda_R^{\text{Japan}} = 1.48 \times 10^5 \\ \lambda_O = 0.05 \end{cases}$$



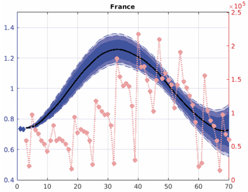
MLE through SAEM

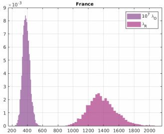


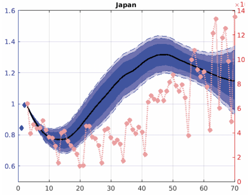




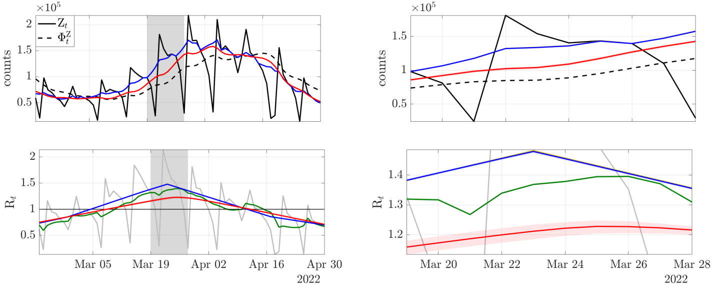
Full Bayesian with flat priors







Full Bayesian with Gamma priors



$\hat{\theta}$ via MLE (no λ_R, λ_O); $\hat{\theta}$ via "state of the art" (no λ_R, λ_O);
$\hat{\theta}$ via MAP (fixed λ_R, λ_O); $\hat{\theta}$ via mean a posteriori (fixed λ_R, λ_O);
$\hat{\theta}$ via full Bayesian and mean a posteriori (integrated w.r.t. λ_R, λ_O)

I. Reproduction number of the Covid19
Extensions (in progress or to come)

Conclusions of Part I.

- Model

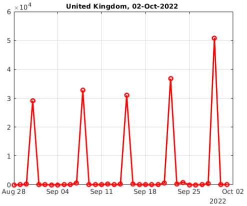
- Poisson: $\mathcal{P}(\tau) \longrightarrow \alpha\, \mathcal{P}(\tau/\alpha)$; or Gamma $\rightsquigarrow$ better fit real counts intrinsic variance
- (few) Missing data $\longrightarrow$ mixture models
- (few) Data: the data are seen as aggregated values

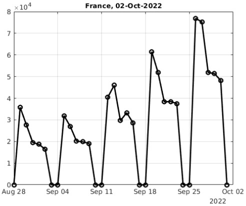
- Calibration of models, with experts

A. Cory (Imperial College, London), A. Flahault (Institut de Santé Globale, Geneva)
- choice of the parameters ϕ
- many ideas for the hyperparameters $(\lambda_{\mathrm{R}}, \lambda_{\mathrm{O}}) \to$ knowledge of experts

- Computational tools for sampling the distributions

- MCMC
- Variational inference





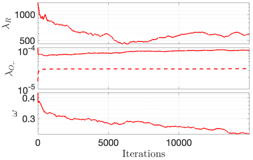
Outliers mixture model

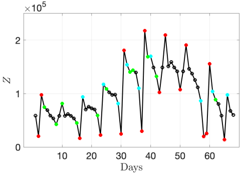
From Abry et al Pascal et al (2023)

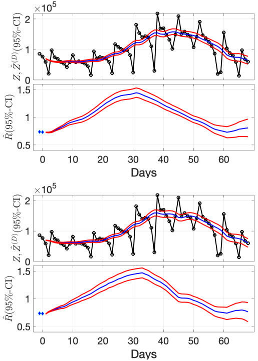
- High variability in daily counts
  ⇝ Two distinct types of counting error to encompass a wider range of behaviors

$$\begin{cases} \mathsf{B}_t \sim \mathcal{B}(\omega) \\ \mathsf{O}_t \,|\, \mathsf{B}_t \sim \text{Laplace}\,(\lambda_{\mathsf{O},1})\mathbb{1}_{\mathsf{B}_t=1} + \text{Laplace}\,(\lambda_{\mathsf{O},2})\mathbb{1}_{\mathsf{B}_t=0} \end{cases} \quad \text{where} \quad \omega \in [0,1]$$

- The rest (Law of Z_t, R_t, etc.) remains unchanged
- Estimation of the R_t's when $(\lambda_\mathsf{R}, \lambda_{\mathsf{O},1}, \lambda_{\mathsf{O},2})$ maximize the likelihood (using SAEM)







Figures (from left to right): - $\hat{\lambda}^k_\mathsf{R}$, $\hat{\lambda}^k_{\mathsf{O},1}$ (solid), $\hat{\lambda}^k_{\mathsf{O},2}$ (dashed) and $\hat{\omega}^k$
- Counts Z_t colored according to the proba. that $|\mathsf{O}_t|$ is large (i.e. $\mathsf{B}_t = 0$)
- `top`: Mixture model, `bottom`: Usual model

II. Estimation of the R_t's via MCMC

For the application : sample from the density $\theta \mapsto \pi_{\mathsf{norm}}(\theta | Z_{1:T}, \lambda_{\mathsf{R}}, \lambda_{\mathsf{O}})$, with $-\ln \pi_{\mathsf{norm}}$ given by up to an additive ctt

$$(\mathsf{R}_{1:T}, \mathsf{O}_{1:T}) \mapsto \sum_{t=1}^{T} \{(\mathsf{R}_t \Phi_t + \mathsf{O}_t) - Z_t \ln(\mathsf{R}_t \Phi_t + \mathsf{O}_t)\} + \lambda_{\mathsf{R}} \|\mathsf{D}\mathsf{R}_{1:T} + \delta\|_1 + \lambda_{\mathsf{O}} \|\mathsf{O}_{1:T}\|_1$$

on the set $\mathcal{D}$.

Forget the application and generalize the problem

The target distribution π_{norm} is known up to a normalizing constant

$$-\ln \pi(\theta) = \begin{cases} f(\theta) + h(\mathsf{A}\theta) & \text{on } \mathcal{D} \\ +\infty & \text{otherwise} \end{cases}$$

- a $\log$-concave density π
- a composite log-density: sum of a
  - convex C^1 function f
  - a proper lower-semicontinuous convex function h composed with a linear operator A
- a domain $\mathcal{D} \subsetneq \mathbb{R}^{2T}$

No closed form expressions for statistics (quantiles, expectation, $\cdots$) $\rightsquigarrow$ Markov Chain Monte Carlo samplers produce points $\{\theta^n, n \geq 0\}$ s.t.

- Approximation of the expectation

$$\int \theta \, \frac{\pi(\theta)}{\int \pi(\tau) \mathrm{d}\tau} \, \mathrm{d}\theta \approx \frac{1}{N} \sum_{n=1}^{N} \theta^n$$

- Approximation of the quantile of order α

$$\theta^{(\lceil \alpha N \rceil, N)}$$

Which MCMC sampler ? Hastings-Metropolis family

- Repeat: starting from $\theta^0 \in \mathbb{R}^{2T}$

- sample a candidate $\theta^{n+\frac{1}{2}} \sim q(\cdot|\theta^n)$
- [AR step] accept this candidate: $\theta^{n+1} = \theta^{n+\frac{1}{2}}$ with probability

$$1 \wedge \frac{\pi(\theta^{n+\frac{1}{2}})}{\pi(\theta^n)} \frac{q(\theta^n|\theta^{n+\frac{1}{2}})}{q(\theta^{n+\frac{1}{2}}|\theta^n)}$$

and otherwise reject: $\theta^{n+1} = \theta^n$

- Popular proposal kernels:
  - Random walk: $q(\cdot|\theta^n) = q(\cdot - \theta^n)$ ex. $\mathcal{N}(\theta^n, \mathsf{C})$
  - Langevin: $q(\cdot|\theta^n) \equiv \theta^n + \gamma \ \nabla \log \pi(\theta^n) + \sqrt{2\gamma}\ \mathcal{N}(0, \mathrm{I})$

For our application

Let us compare
a Gaussian random walk approach

$$q(\cdot|\theta^n) \sim \theta^n + \sqrt{2\gamma}\,\mathcal{N}(0, \mathrm{I})$$

with Langevin-based approaches

$$q(\cdot|\theta^n) \sim \mu_\gamma(\theta^n) + \sqrt{2\gamma}\,\mathcal{N}(0, \mathrm{I})$$

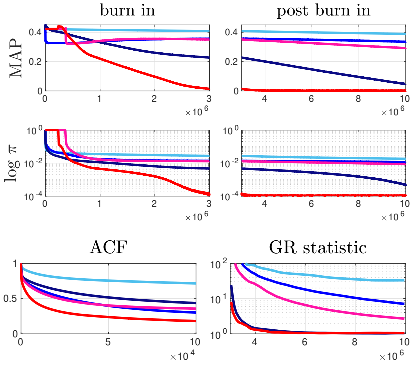
where the drift μ_γ uses first order informations on $\log \pi$.

MAP and log π: how fast the chain $\{\theta^n, n \geq 0\}$ moves to the maximizer θ^{MAP}; and $\log \pi(\theta^n)$ moves to $\log \pi(\theta^{\mathsf{MAP}})$.
Good values of the criteria: zero.

ACF: Related to the variance in a CLT.
Good value of the criterion: zero.

Gelman Rubin statistic: how fast the chain forgets its initial value.

Good value of the criterion: one.



Random Walk with $\mathrm{C} = \mathrm{I}$, with another matrix C_1 and with another matrix C_2
Proximal-Gradient based drift with a covariance matrix C and with another one $\tilde{\mathrm{C}}$

Langevin-based approaches in the literature

$$-\log \pi(\cdot) = f(\cdot) + h(\mathsf{A}\cdot) \qquad \text{on } \mathcal{D}$$

- which drift function $\theta \mapsto \mu_\gamma(\theta)$?

- for smooth $\log \pi$: $\mu_\gamma(\theta) = \theta + \gamma \nabla \log \pi(\theta)$

- for non-smooth $\log \pi$: $\mu_\gamma(\theta) = \theta - \gamma \nabla f(\theta) - \gamma \nabla \mathsf{M}^\gamma_{h(\mathsf{A}\cdot)}(\theta)$

the gdt of the Moreau envelope of $h(\mathsf{A}\cdot)$ has no closed form expression in our case

- with or without the AR step ?

- with the AR step: target π

for managing the domain $\mathcal{D}$ in our case, $\mathcal{D} \subsetneq \mathbb{R}^{2T}$

the ratio of densities in $\mathbb{R}^{2T}$ in our case, T is large

- without the AR step: target π_γ and control the asymptotic bias $\text{dist}(\pi, \pi_\gamma)$

how to manage the domain $\mathcal{D}$ The "Moreau envelope" approach was proposed in the case $\mathcal{D} = \mathbb{R}^{2T}$

Which first order informations on $\log \pi$ are available in our case ?

$$-\log \pi(\cdot) = f(\cdot) + h(\mathsf{A}\cdot) \qquad \text{on } \mathcal{D}$$

- the gradient of f
- the subgradient of $h(\cdot)$ and the subgradient of $h(\mathsf{A}\cdot)$
- the proximal of f $\qquad \mathrm{prox}_{\gamma f}(\tau) := \mathrm{argmin}_\theta \left(\gamma f(\theta) + \frac{1}{2}\|\theta - \tau\|^2\right)$
- the proximal of h $\qquad$ but **NOT** the proximal of $h(\mathsf{A}\cdot)$

and, when A is invertible, the lemma :

If
$\{\tilde{\theta}^n, n \geq 0\}$ is a Markov chain with invariant distribution $\propto \pi(\mathsf{A}^{-1}\cdot)$,
then
$\{\mathsf{A}^{-1}\tilde{\theta}^n, n \geq 0\}$ is a Markov chain with invariant distribution $\pi_{\mathrm{norm}}(\cdot)$

$p = \mathrm{prox}_{\gamma f}(\tau)$ iff $\tau \in p + \gamma\, \partial f(p)$ i.e. $\tau - p \in \gamma\, \partial f(p)$

Our contribution: many ideas for the drift μ_γ

- sampling in the original space: $q(\cdot|\theta) = \mu_\gamma(\theta) + \sqrt{2\gamma}\,\mathcal{N}(0, \mathrm{I})$

- Full sub-gradient

$$\mu_\gamma(\theta) := \theta - \gamma \nabla f(\theta) - \gamma A^\top H(\mathsf{A}\theta) \qquad H(u) \in \partial h(u)$$

- Proximal and sub-gradient

$$\mu_\gamma(\theta) := \mathrm{prox}_{\gamma f}\left(\theta - \gamma A^\top H(\mathsf{A}\theta)\right)$$

- with a change of geometry: $q(\cdot|\theta) = \mu_\gamma(\theta) + \sqrt{2\gamma}\,\mathcal{N}(0, \mathsf{A}^{-1}\mathsf{A}^{-\top})$

- Full sub-gradient

Rmk: Tempered Langevin

$$\mu_\gamma(\theta) := \theta - \gamma \mathsf{A}^{-1}\mathsf{A}^{-\top} \nabla f(\theta) - \gamma \mathsf{A}^{-1} H(\mathsf{A}\theta) \qquad H(u) \in \partial h(u)$$

- Sub-gradient and proximal

$$\mu_\gamma(\theta) := \mathrm{prox}_{\gamma h}\left(\mathsf{A}\theta - \gamma \mathsf{A}^{-\top} \nabla f(\theta)\right)$$

Comparison by numerical criteria (1/2)

Chevallier and Fort (2024)

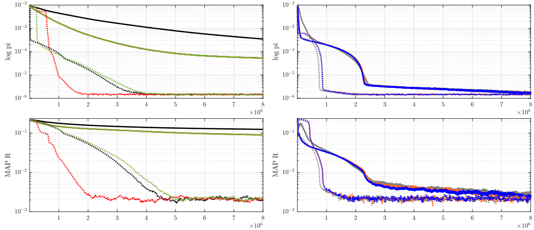


Figure: $\log \pi$ (top) and MAP-R (bottom) criteria.

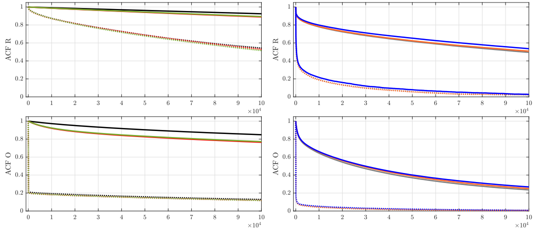


Figure: ACF criteria for $R_{1:T}$ (top) and $O_{1:T}$ (bottom)

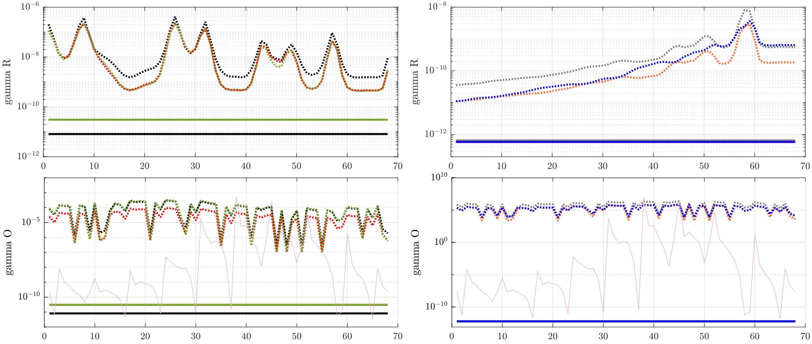
(left) Original space: RW, FSG, Prox-SG

(right) After a change of geometry: inv-RW, inv-FSG, SG-Prox.

No Gibbs samplers are in solid lines and one-at-a-time Gibbs samplers are in dotted lines.

Comparison by numerical criteria (2/2)

Chevallier and Fort (2024)



The step size γ, for algorithms run in the original space (left, RW, FSG, Prox-SG) or algorithms after a change of geometry (right, inv-RW, inv-FSG, SG-Prox).
`No Gibbs` samplers are in solid lines and `one-at-a-time Gibbs` samplers are in dotted lines.
Bottom row: $Z_1, \cdots, Z_T$ are displayed in light pink on an independent y-scale.

Conclusions of Part II: next steps

- Methodological:

- from primal dual optimization methods to new drift functions μ_γ.

- Numerical:

- comparison on a toy example
- role of (λ_R, λ_O)
- another application (statistical signal or image processing)

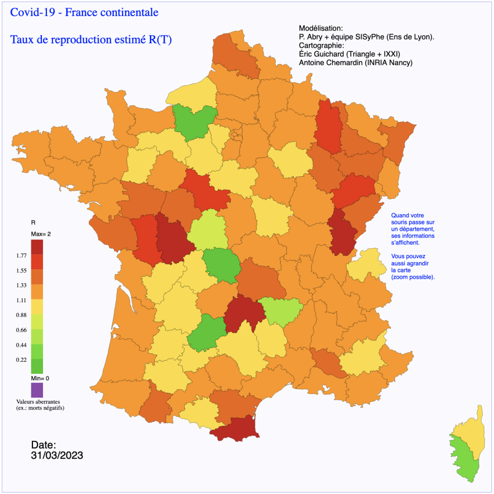
- Theoretical:

- an approach based on "drift inequalities"
- or an approach based on "discretization of a SDE"
-  SDE on a domain

- Methodological:

- insert the *best* strategy into Sequential Monte Carlo samplers for online estimation.

At the end of OptiMoCSI ⋯ barthes.enssib.fr/coronavirus/cartes/RFrance



carte interactive Monde : http://barthes.enssib.fr/coronavirus/cartes/dev/monde/monde.html

Appendix

Maximum likelihood estimation of λ_R and λ_O through EM algorithm

From Abry et al Pascal et al (2023)

- Maximum likelihood criterion: $\Lambda = (\lambda_R, \lambda_O)$

$$\underset{\Lambda>0}{\operatorname{argmax}}\ \pi\left(Z_{1:T}\,;\,\Lambda\right) \;=\; \underset{\Lambda>0}{\operatorname{argmax}}\ \int_{\mathcal{D}} \pi\left(Z_{1:T},\theta\,;\,\Lambda\right) \mathrm{d}\theta$$

- **Expectation-Maximization** (EM) algorithm:

E-step: Compute the conditional expected log-likelihood

$$Q(\Lambda\,|\,\Lambda^k) = \int_{\mathcal{D}} \ln \pi\left(Z_{1:T},\theta\,;\,\Lambda\right)\, \pi\big(\theta\,\big|\,Z_{1:T}\,;\,\Lambda^k\big)\,\mathrm{d}\theta$$

M-step: Maximize $Q(\cdot\,|\,\Lambda^k)$ in the feasible set $\{\Lambda > 0\}$

Maximum likelihood estimation of λ_R and λ_O through EM algorithm

From Abry et al Pascal et al (2023)

- Maximum likelihood criterion: $\Lambda = (\lambda_R, \lambda_O)$

$$\underset{\Lambda>0}{\operatorname{argmax}} \ \pi\left(Z_{1:T}\,;\,\Lambda\right) = \underset{\Lambda>0}{\operatorname{argmax}} \int_{\mathcal{D}} \pi\left(Z_{1:T}, \theta\,;\,\Lambda\right) \mathrm{d}\theta$$

- **Expectation-Maximization** (EM) algorithm:

E-step: Compute the conditional expected log-likelihood

$$Q(\Lambda \,|\, \Lambda^k) = cste + T \ln \lambda_R - \lambda_R\, S_R(\Lambda^k) + T \ln \lambda_O - \lambda_O\, S_O(\Lambda^k)$$

$$S_R(\Lambda^k) := \int_{\mathcal{D}} \|\mathrm{D}R_{1:T}\|_1\, \pi\big(\theta \,\big|\, Z_{1:T}\,;\,\Lambda^k\big)\, \mathrm{d}\theta \qquad S_O(\Lambda^k) := \int_{\mathcal{D}} \|O_{1:T}\|_1\, \pi\big(\theta \,\big|\, Z_{1:T}\,;\,\Lambda^k\big)\, \mathrm{d}\theta$$

M-step: Maximize $Q(\cdot \,|\, \Lambda^k)$ in the feasible set $\{\Lambda > 0\}$

$$\lambda_R^{k+1} := \frac{T}{S_R(\Lambda^k)} \quad \text{and} \quad \lambda_O^{k+1} := \frac{T}{S_O(\Lambda^k)}$$

Maximum likelihood estimation of λ_R and λ_O through EM algorithm

From Abry et al Pascal et al (2023)

- Maximum likelihood criterion: $\Lambda = (\lambda_R, \lambda_O)$

$$\underset{\Lambda>0}{\operatorname{argmax}}\ \pi\left(Z_{1:T}\,;\,\Lambda\right) = \underset{\Lambda>0}{\operatorname{argmax}} \int_{\mathcal{D}} \pi\left(Z_{1:T}, \theta\,;\,\Lambda\right) \mathrm{d}\theta$$

- **Expectation-Maximization** (EM) algorithm: $\rightsquigarrow$ Stochastic Approximation EM

E-step: Compute the conditional expected log-likelihood

$$Q(\Lambda \,|\, \Lambda^k) = cste + T \ln \lambda_R - \lambda_R\, S_R(\Lambda^k) + T \ln \lambda_O - \lambda_O\, S_O(\Lambda^k)$$

$$S_R(\Lambda^k) := \int_{\mathcal{D}} \|\mathrm{D}R_{1:T}\|_1\, \pi\big(\theta \,\big|\, Z_{1:T}\,;\,\Lambda^k\big)\, \mathrm{d}\theta \qquad S_O(\Lambda^k) := \int_{\mathcal{D}} \|O_{1:T}\|_1\, \pi\big(\theta \,\big|\, Z_{1:T}\,;\,\Lambda^k\big)\, \mathrm{d}\theta$$

M-step: Maximize $Q(\cdot \,|\, \Lambda^k)$ in the feasible set $\{\Lambda > 0\}$

$$\lambda_R^{k+1} := \frac{T}{S_R(\Lambda^k)} \quad \text{and} \quad \lambda_O^{k+1} := \frac{T}{S_O(\Lambda^k)}$$

Maximum likelihood estimation of λ_R and λ_O through SAEM algorithm

From Abry et al Pascal et al (2023)

- **Stochastic Approximation Expectation-Maximization** (SAEM) algorithm:

S-step: - Generate realizations of the latent variable θ under the conditional density $\pi\big(\,\cdot\,\big|\,Z_{1:T}\,;\,\Lambda^k\big)$

- Construct $\widehat{Q}^k$ Monte Carlo approximation of $Q(\cdot\,|\,\Lambda^k)$

SA-step: Given an initial approximation Q^0 of $Q(\cdot\,|\,\Lambda^k)$

$$Q^{k+1}(\Lambda) = Q^k(\Lambda) + \gamma^k \left(\widehat{Q}^k(\Lambda) - Q^k(\Lambda)\right)$$

M-step: Maximize Q^k in the feasible set $\{\Lambda > 0\}$

$$\lambda_R^{k+1} := \frac{T}{S_R^{k+1}} \qquad \text{and} \qquad \lambda_O^{k+1} := \frac{T}{S_O^{k+1}}$$

Maximum likelihood estimation of λ_R and λ_O through SAEM algorithm

From Abry et al Pascal et al (2023)

- **Stochastic Approximation Expectation-Maximization** (SAEM) algorithm:

S-step: - Generate realizations of the latent variable θ under the conditional density $\pi\left(\cdot \;\middle|\; Z_{1:T}\,;\,\Lambda^k \right)$

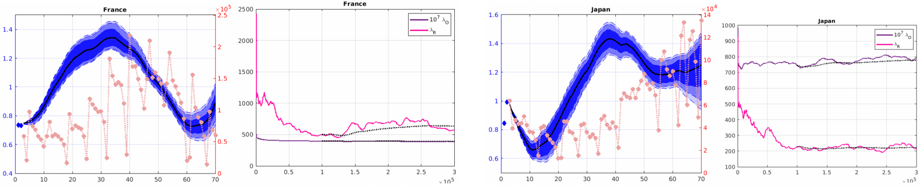
- Construct $\widehat{S}_R^k$ and $\widehat{S}_O^k$ Monte Carlo approx. of $S_R(\Lambda^k)$ and $S_O(\Lambda^k)$

SA-step: Given initial approximations S_R^0 and S_O^0,

$$S_R^{k+1} = S_R^k + \gamma_R^k \left(\widehat{S}_R^k - S_R^k \right) \quad \text{and} \quad S_O^{k+1} = S_O^k + \gamma_O^k \left(\widehat{S}_O^k - S_O^k \right)$$

M-step: Maximize Q^k in the feasible set $\{\Lambda > 0\}$

$$\lambda_R^{k+1} := \frac{T}{S_R^{k+1}} \quad \text{and} \quad \lambda_O^{k+1} := \frac{T}{S_O^{k+1}}$$



Full Bayesian model: Overcoming λ_R and λ_O dependency

From Abry et al Pascal et al (2023)

- **Idea**: - Consider $\Lambda = (\lambda_R, \lambda_O)$ as a random variable
  - Estimate the R_t's by integration w.r.t Λ

- Flat prior on Λ
- Hence: Joint distribution $\pi(\theta, \Lambda \,|\, Z_{1:T})$ proportional to the posterior $\pi(\theta \,|\, Z_{1:T}\,;\, \Lambda)$, *i.e:*

$$\pi(\theta, \Lambda \,|\, Z_{1:T}) \propto \pi(\theta \,|\, Z_{1:T}\,;\, \Lambda)$$

- Point estimates and Credibility interval of R_t's:
  MCMC approximation of $\pi(\theta \,|\, Z_{1:T}\,;\, \Lambda)$