Pandemic Intensity Estimation from Stochastic Approximation-based Algorithms

Patrice Abry, Juliette Chevallier, Gersende Fort, Barbara Pascal

Pandemic intensity estimation from Stochastic Approximation-based algorithms

Patrice Abry
Laboratoire de Physique
CNRS, ENS Lyon, FR
patrice.abry@ens-lyon.fr

Juliette Chevallier
IMT, UMR 5219, INSA Toulouse
Université de Toulouse, FR

Gersende Fort
IMT, UMR 5219, CNRS
Université de Toulouse, FR

Barbara Pascal
Nantes Université, ECN
CNRS, LS2N, FR
barbara.pascal@cnrs.fr

firstname.lastname@math.univ-toulouse.fr

***Abstract*—Pandemic intensity monitoring, from the earliest stages of the pandemic outbreak, constitutes a critical scientific challenge with major societal stakes. The task is significantly complicated by the low quality of reported infection counts, stemming from emergency and crisis contexts, and by the need for regular (daily) updates, while the pandemic is still active. The present work first proposes a parametric Hidden Markov Model (HMM) aiming to account jointly for epidemic propagation mechanisms and for low-quality data, while imposing epidemic-compliant constraints on the time-varying reproduction number, considered as a proxy for pandemic intensity quantification. Second, and to avoid the arbitrary or expert-based tuning of the parameters of the HMM, data-driven automated selection procedures are devised relying on tailoring a stochastic Expectation-Maximization algorithm. Credibility interval-based estimation of the time-varying reproduction number, modeled as a hidden variable, is then obtained from Monte Carlo sampling. The potential of the tools devised here is illustrated on real Covid19 daily new infection counts from Johns Hopkins University repository.**

***Index Terms*—Statistical modelization, Latent variable models, Statistical inference, Stochastic Expectation-Maximization, Covid19 pandemic, Reproduction number.**

I. Introduction

Context. With the Covid19 *event*, monitoring a pandemic intensity in the active phase has been recognized as a critical challenge with high societal stakes [1], and has drawn significant research efforts [2]. Retrospective assessment of the intensity of a pandemic once it has ended, is generally carried out using *compartmental models* [3], [4]. However, these models suffer from high computational costs or demanding parameterization, inducing poor robustness to the low-quality daily counts of newly reported infections, the key ingredient for assessing pandemic intensity. Because of the need to collect test outcomes in emergency and centralized contexts, Covid19 counts reported by national health agencies were, for all countries, massively corrupted with essentially two classes of atypical values (see Fig. 1[right]): *(i)* non-reported counts (week-ends, non-working days, thus inducing a weekly pseudo-seasonality) and/or *(ii)* inaccurate counts. Data corruption varied from one country to another and for the same country between the different phases of the pandemic.

Work partly supported by the Foundation *Simone et Cino Del Duca* under the project OpSiMorE, and by the French National Research Agency under projects ANR-19-CE23-0017-01 MASDOL and ANR-23-CE48 OptiMoCSI.

Related works. Strategies for *within pandemic* intensity monitoring were devised, often focused on a time-varying reproduction number, R_t, that quantifies (dynamically) the mean number of second infections stemming from one same primary infection [5]–[8]. To account for Covid19 data limited quality, the epidemic model in [8] was recast as a functional optimization problem favoring piecewise linear temporal evolution of R_t and assuming mild sparsity of atypical values, without using any calendar information [9], [10]. When sanitary or economic countermeasures must be designed, confidence assessment is critical to decision-makers. Seeing R_t as a random variable and modeling the counts as the observations in a parametric Hidden Markov Model (HMM; see e.g. [11]), Markov Chain Monte Carlo (MCMC) sampling strategies were further devised to produce credibility interval-based (CI) estimates of R_t [12], [13]. Though promising, these contributions suffer from major limitations, addressed here: *(i)* the possibly multiple causes of atypical counts are not modeled with sufficient versatility; *(ii)* parameters of the HMM were so far tuned by experts. In this paper, we use a Maximum Likelihood approach to estimate the parameters, and derive a computation via a Stochastic Approximation algorithm [14].

Goals, contributions, and outline. The present work aims to devise strategies permitting a robust, accurate, parameter-free, CI-based assessment of the time evolution of the intensity of an epidemic, from daily new infection counts possibly highly corrupted in several ways. As a first contribution, a HMM is constructed by combining epidemic mechanisms, temporal evolution regularity constraints, and count miss-report modeling based on generic sparsity arguments and designed to account for the multiple natures and sizes of erroneous counts (Section II-A). This is complemented by devising MCMC sampling strategies that output CI-based estimations of R_t and, as a side result, of *denoised* realistic daily infection counts (Section II-B). As a second contribution, a Stochastic Approximation version of Expectation-Maximization (EM) [15] is devised: it permits automated data-driven parameter calibration, thus avoiding arbitrariness in expert-knowledge a priori selection or a posteriori validation (Section II-C). Finally, the potential and performance of the statistical estimation tools devised here are shown in action and assessed on real Covid19 data extracted from Johns Hopkins University repository, https://coronavirus.jhu.edu/ (Section III).

II. Pandemic reproduction number estimation

A. A parametric HMM

The present model aims to establish a link between the observed daily infection count Z_t and the non-observed daily reproduction number R_t. Data corruption is encapsulated under the general term of O_t which is unknown by construction. The first contribution of this paper is to propose a mixture model for the erroneous counts O_t, in order to account for small and large errors due to strongly misreported counts or even missing reports. As in [12], we propose a model in which $\{\mathsf{Z}_t, \mathsf{R}_t, \mathsf{O}_t, t \in \mathbb{Z}\}$ are a HMM, taking values in $\mathbb{Z} \times \mathbb{R} \times \mathbb{R}$.

Given the positive *serial interval function* $\mathbf{\Phi} := (\Phi_u)_{1 \leqslant u \leqslant \tau_\phi}$ describing the average infectiousness profile after infection (see [8]), set $\Phi_t^{\mathbf{Z}} := \sum_{u=1}^{\tau_\phi} \Phi_u \mathsf{Z}_{t-u}$ for any $t \in \mathbb{Z}$. Conditionally to the past, and more precisely to $(\mathsf{R}_t, \mathsf{O}_t, \Phi_t^{\mathbf{Z}})$, the new infections count Z_t at day t is modeled as

$$\begin{cases} \mathcal{P}\big(\mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t\big) & \text{if } \mathsf{R}_t \geqslant 0 \quad \text{and} \quad \mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t > 0, \\ \delta_0 & \text{if } \mathsf{R}_t \geqslant 0 \quad \text{and} \quad \mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t = 0, \\ \nu & \text{if } \mathsf{R}_t < 0 \quad \text{or} \quad \mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t < 0, \end{cases} \tag{1}$$

where $\mathcal{P}(\mathcal{I})$ denotes the Poisson distribution with parameter $\mathcal{I}$, ν is any distribution over the negative integers $\{-1, -2, \ldots\}$, and δ_0 is the Dirac mass at zero. The first two cases are introduced in [8], and the third case in [12] to define a rigorous statistical model. Since the counts Z_t are non-negative, such a model implies that: *(i)* a negative reproduction number or a negative Poisson parameter never occurs; *(ii)* the conditional distribution of Z_t given $(\mathsf{R}_t, \mathsf{O}_t, \Phi_t^{\mathbf{Z}})$ is

$$p\left(\mathsf{Z}_t | \mathsf{R}_t, \mathsf{O}_t, \Phi_t^{\mathbf{Z}}\right) := \frac{(\mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t)^{\mathsf{Z}_t}}{\mathsf{Z}_t!}\, \mathrm{e}^{-(\mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t)}, \tag{2}$$

if $(\mathsf{R}_t, \mathsf{O}_t) \in \mathcal{D}_t$, where $\mathcal{D}_t := \{\mathsf{R}_t \geqslant 0 \text{ and } \mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t \geqslant 0\}$; by convention, $0^0 = 1$. Following [12], we assume that R_t and O_t are independent conditionally to the past. For the R_t's,

$$p\left(\mathsf{R}_t | \mathsf{R}_{t-1}, \mathsf{R}_{t-2}; \lambda_{\mathsf{R}}\right) := \frac{\lambda_{\mathsf{R}}}{8}\, \mathrm{e}^{-\frac{\lambda_{\mathsf{R}}}{4} |\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2}|}; \tag{3}$$

it implies that the mode of the joint distribution $p(\mathsf{R}_1, \cdots, \mathsf{R}_T | \mathsf{R}_0, \mathsf{R}_{-1}; \lambda_{\mathsf{R}})$ is sparse, favoring a sparse second derivative of the function $t \mapsto \mathsf{R}_t$'s and thus a piecewise linear evolution of the function [10]. Regarding the errors O_t, several modeling choices are available to us. In [12], they are assumed to be independent and modeled using a single-type Laplace distribution. Nevertheless, as shown for example in Fig. 1[right], there may be a high variability in daily counts. Here, we propose to allow two distinct types of counting errors to encompass a broader range of behaviors. This duality is modeled using a mixture of two Laplace distributions both centered on zero, but with respective scale parameters $\lambda_{\mathsf{O},1}$ and $\lambda_{\mathsf{O},2}$. Given a $\{0,1\}$-valued Bernoulli variable B_t coding for the error type, we set

$$p\left(\mathsf{O}_t | \mathsf{B}_t; \lambda_{\mathsf{O},1}, \lambda_{\mathsf{O},2}\right) := \left(\frac{\lambda_{\mathsf{O},1}}{2}\, \mathrm{e}^{-\lambda_{\mathsf{O},1} |\mathsf{O}_t|}\right)^{\mathsf{B}_t} \times \left(\frac{\lambda_{\mathsf{O},2}}{2}\, \mathrm{e}^{-\lambda_{\mathsf{O},2} |\mathsf{O}_t|}\right)^{1-\mathsf{B}_t}. \tag{4}$$

The Bernoulli variables are independent and identically distributed with success parameter $\omega \in [0,1]$

$$p(\mathsf{B}_t; \omega) := \omega^{\mathsf{B}_t} (1-\omega)^{1-\mathsf{B}_t}. \tag{5}$$

The model (4)-(5) encompasses the single-mode model developed in [12] by setting $\omega = 1$. Eqs. (1)-(5) define a HMM, parameterized by $\Lambda := (\lambda_{\mathsf{R}}, \lambda_{\mathsf{O},1}, \lambda_{\mathsf{O},2}, \omega)$ in $\mathcal{E} := (\mathbb{R}_{>0})^3 \times [0,1]$.

B. Estimation of the daily reproduction numbers R_t

The estimation of R_t is based on the a posteriori distribution of $\mathbf{R} := (\mathsf{R}_1, \ldots, \mathsf{R}_T)$ given the observations $\mathbf{Z} := (\mathsf{Z}_1, \ldots, \mathsf{Z}_T)$ and the initial values $\mathsf{I} := \big(\mathsf{R}_{-1}, \mathsf{R}_0, \mathsf{Z}_{-\tau_\phi+1}, \ldots, \mathsf{Z}_0\big)$. We set $\mathbf{O} := (\mathsf{O}_1, \ldots, \mathsf{O}_T)$ and $\mathbf{B} := (\mathsf{B}_1, \ldots, \mathsf{B}_T)$. Point estimations and CIs for $\mathsf{R}_1, \cdots, \mathsf{R}_T$ may rely on the expectation, the median and more generally on the quantiles of this conditional distribution. Numerically, these statistics are estimated from a MCMC approximation of $\pi(\cdot|\mathbf{Z}, \mathsf{I}; \Lambda)$, the a posteriori distribution of $(\mathbf{R}, \mathbf{O}, \mathbf{B})$ given $\mathbf{Z}$ and I when the parameters of the HMM are equal to Λ.

The a posteriori distribution. From (1)-(5), we have that $\pi(\cdot|\mathbf{Z}, \mathsf{I}; \Lambda)$ is proportional to the joint distribution of $(\mathbf{R}, \mathbf{O}, \mathbf{B}, \mathbf{Z})$ which is $\exp(U_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}, \mathbf{B}; \Lambda))$ on $\mathcal{D} := \bigcap_{t=1}^T \mathcal{D}_t$ and 0 otherwise, with

$$\begin{aligned} U_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}, \mathbf{B}; \Lambda) &:= F_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}) + T \ln \lambda_{\mathsf{R}} \\ &- \lambda_{\mathsf{R}}\, S_{\mathsf{R}}(\mathsf{R}_{-1}, \mathsf{R}_0, \mathbf{R}) - \lambda_{\mathsf{O},1}\, S_{\mathsf{O},1}(\mathbf{O}, \mathbf{B}) - \lambda_{\mathsf{O},2}\, S_{\mathsf{O},2}(\mathbf{O}, \mathbf{B}) \\ &+ S_{\mathsf{B},1}(\mathbf{B})\, \ln(\omega\, \lambda_{\mathsf{O},1}) + S_{\mathsf{B},2}(\mathbf{B})\, \ln((1-\omega)\, \lambda_{\mathsf{O},2}); \end{aligned}$$

$$F_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}) := \sum_{t=1}^T \left(\mathsf{Z}_t \ln(\mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t) - (\mathsf{R}_t\, \Phi_t^{\mathbf{Z}} + \mathsf{O}_t)\right),$$

$$S_{\mathsf{R}}(\mathsf{R}_{-1}, \mathsf{R}_0, \mathbf{R}) := \frac{1}{4} \sum_{t=1}^T |\mathsf{R}_t - 2\mathsf{R}_{t-1} + \mathsf{R}_{t-2}|,$$

$$S_{\mathsf{O},1}(\mathbf{O}, \mathbf{B}) := \sum_{t=1}^T \mathsf{B}_t |\mathsf{O}_t|, \quad S_{\mathsf{O},2}(\mathbf{O}, \mathbf{B}) := \sum_{t=1}^T (1-\mathsf{B}_t) |\mathsf{O}_t|,$$

$$S_{\mathsf{B},1}(\mathbf{B}) := \sum_{t=1}^T \mathsf{B}_t, \quad S_{\mathsf{B},2}(\mathbf{B}) := \sum_{t=1}^T (1-\mathsf{B}_t) = T - S_{\mathsf{B},1}(\mathbf{B}).$$

$\pi(\cdot|\mathbf{Z}, \mathsf{I}; \Lambda)$ has an intricate expression and is known up to a normalizing constant; thus a MCMC approximation is used.

The `Gibbs-PGdual` MCMC sampler. The approximation of $\pi(\cdot|\mathbf{Z}, \mathsf{I}; \Lambda)$ is obtained by running a Gibbs sampler. Let us derive the conditional distributions associated with this joint distribution. Conditionally to $(\mathbf{R}, \mathbf{O})$, $\mathbf{B}$ has independent components with a Bernoulli distribution such that $\mathsf{B}_t = 1$ with probability

$$\left(1 + {}^{(1-\omega)\lambda_{\mathsf{O},2} \exp(-\lambda_{\mathsf{O},2} |\mathsf{O}_t|)}\big/_{\omega \lambda_{\mathsf{O},1} \exp(-\lambda_{\mathsf{O},1} |\mathsf{O}_t|)}\right)^{-1}.$$

Given $(\mathbf{O}, \mathbf{B})$, the log-density of $\mathbf{R}$ is, up to an additive constant, $\mathbf{R} \mapsto F_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}) - \lambda_{\mathsf{R}} S_{\mathsf{R}}(\mathsf{R}_{-1}, \mathsf{R}_0, \mathbf{R})$ on the set $\mathcal{D}$ and $-\infty$ otherwise. Finally, given $(\mathbf{R}, \mathbf{B})$, the log-density of $\mathbf{O}$ is, up to an additive constant, $\mathbf{O} \mapsto F_{\mathbf{Z},\mathsf{I}}(\mathbf{R}, \mathbf{O}) - \lambda_{\mathsf{O},1} S_{\mathsf{O},1}(\mathbf{O}, \mathbf{B}) - \lambda_{\mathsf{O},2} S_{\mathsf{O},2}(\mathbf{O}, \mathbf{B})$ on the set $\mathcal{D}$ and $-\infty$ otherwise. Exact sampling from the conditional distributions of

R and O is not possible. We replace such a sampling with one iteration of the `PGdual` kernel (see [12, Section III-C]). This kernel was designed for log-target densities being the sum of a continuously differentiable function (here, the $F_{\mathsf{Z},\mathsf{I}}$ function) and a prox-friendly function (here, the $S_{\mathsf{O},\ell}$ functions) possibly combined with a linear operator (here, the S_{R} function) when a support exists (here, the set $\mathcal{D}$). `PGdual` was successfully applied to a Bayesian analysis of the Covid19 reproduction number (see e.g. [13], [16], [17]).

C. Data-driven calibration of the parametric HMM

The parameters Λ are unknown. In this work, we propose to compute the Maximum-Likelihood estimator of the observations Z conditionally to the initial values I:

$$\widehat{\Lambda} \in \operatorname*{argmax}_{\Lambda\in\mathcal{E}} \ln \mathcal{L}(\mathsf{Z}|\mathsf{I};\Lambda). \tag{6}$$

The likelihood $\mathcal{L}$ is defined through an integral over the hidden variables (see Section II-A) and does not have an explicit expression. The second contribution of our paper is to derive a stochastic optimization procedure to solve (6). It is based on Expectation-Maximization (EM) [15]. Yet, the `E-step` of EM is not tractable, and we propose to tackle this intractability by using a Stochastic Approximation method ([14], [18], [19]).

The EM algorithm. From Section II-A, $\mathcal{L}(\mathsf{Z}|\mathsf{I};\Lambda)$ is equal to

$$\ln \sum_{\mathbf{b}\in\{0,1\}^T} \int_{\mathcal{D}} \exp\left(U_{\mathsf{Z},\mathsf{I}}(\mathbf{r},\mathbf{o},\mathbf{b};\Lambda)\right) \mathrm{d}\mathbf{r}\,\mathrm{d}\mathbf{o},$$

up to an additive constant independent of Λ. Given the current value of the parameters $\Lambda^{(k)}$, the `E-step` of EM consists in computing the function $\mathcal{Q}(\cdot;\Lambda^{(k)})$ defined by

$$\Lambda \mapsto \sum_{\mathbf{b}\in\{0,1\}^T} \int_{\mathcal{D}} U_{\mathsf{Z},\mathsf{I}}(\mathbf{r},\mathbf{o},\mathbf{b};\Lambda)\pi(\mathbf{r},\mathbf{o},\mathbf{b}|\mathsf{Z},\mathsf{I};\Lambda^{(k)})\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{o}.$$

Then, the `M-step` updates the parameters by setting $\Lambda^{(k+1)} \in \operatorname{argmax}_{\mathcal{E}} \mathcal{Q}(\cdot;\Lambda^{(k)})$. In our case, it is readily seen from the expression of U, that the $\mathcal{Q}$-function is fully determined by the computation of the expectations of so-called *sufficient statistics*

$$\bar{\mathsf{S}}_{\mathtt{all}}(\Lambda^{(k)}) := \sum_{\mathbf{b}\in\{0,1\}^T} \int_{\mathcal{D}} S_{\mathtt{all}}(\mathbf{r},\mathbf{o},\mathbf{b})\pi(\mathbf{r},\mathbf{o},\mathbf{b}|\mathsf{Z},\mathsf{I};\Lambda^{(k)})\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{o} \tag{7}$$

where $S_{\mathtt{all}} := (S_{\mathsf{R}}, S_{\mathsf{O},1}, S_{\mathsf{O},2}, S_{\mathsf{B},1}, S_{\mathsf{B},2})$ collects all the sufficient statistics S_x (see Section II-B). On the left-hand side of (7), the dependence upon Z and I is omitted. The `M-step` is explicit and gets into $\Lambda^{(k+1)} := \mathsf{T}(\bar{\mathsf{S}}_{\mathtt{all}}(\Lambda^{(k)}))$, where $\mathsf{T}(s_{\mathsf{R}}, s_{\mathsf{O},1}, s_{\mathsf{O},2}, s_{\mathsf{B},1}, s_{\mathsf{B},2}) := \operatorname{argmax}_{\Lambda\in\mathcal{E}} T\ln\lambda_{\mathsf{R}} - \lambda_{\mathsf{R}} s_{\mathsf{R}} - \sum_{\ell=1}^{2} \lambda_{\mathsf{O},\ell} s_{\mathsf{O},\ell} + s_{\mathsf{B},1}\ln(\omega\lambda_{\mathsf{O},1}) + s_{\mathsf{B},2}\ln((1-\omega)\lambda_{\mathsf{O},2})$, i.e.,

$$\lambda_{\mathsf{R}}^{(k+1)} := \frac{T}{\bar{\mathsf{S}}_{\mathsf{R}}(\Lambda^{(k)})}, \quad \omega^{(k+1)} := \frac{\bar{\mathsf{S}}_{\mathsf{B},1}(\Lambda^{(k)})}{T}, \tag{8}$$

$$\lambda_{\mathsf{O},\ell}^{(k+1)} := \frac{\bar{\mathsf{S}}_{\mathsf{B},\ell}(\Lambda^{(k)})}{\bar{\mathsf{S}}_{\mathsf{O},\ell}(\Lambda^{(k)})}, \qquad \ell\in\{1,2\}. \tag{9}$$

However, due to the intricate expression of π, none of the expectations defining $\bar{\mathsf{S}}_{\mathtt{all}}(\Lambda^{(k)})$ can be exactly computed: hence, EM does not apply.

Stochastic Approximation EM. A stochastic version of EM is used here, namely the *Stochastic Approximation EM* (SAEM) algorithm proposed in [20] (see e.g. [21]–[25] for applications in Signal Processing). It replaces the intractable expectations $\bar{\mathsf{S}}_x(\Lambda)$ in (8)-(9) with a random approximation. More precisely, SAEM defines a (random) sequence of parameters $\{\hat{\Lambda}^{(k+1)}, k\geqslant 0\}$ by setting $\hat{\Lambda}^{(k+1)} := \mathsf{T}(\hat{S}_{\mathtt{all}}^{(k+1)})$, where $\hat{S}_{\mathtt{all}}^{(k+1)}$ is a random approximation of $\bar{\mathsf{S}}_{\mathtt{all}}(\hat{\Lambda}^{(k)})$. This approximation is defined iteratively as follows: Given sequences of learning rates $\{\gamma_x^{(k)}, k\geqslant 1\}$ for $x\in\{\mathsf{R},(\mathsf{O},1),(\mathsf{O},2),(\mathsf{B},1)\}$, SAEM computes

$$\hat{S}_x^{(k+1)} = \hat{S}_x^{(k)} + \gamma_x^{(k+1)}\left(\mathcal{H}_x^{(k+1)} - \hat{S}_x^{(k)}\right),$$

where $\mathcal{H}_x^{(k+1)}$ is a random oracle for $\bar{\mathsf{S}}_x(\hat{\Lambda}^{(k)})$. We produce the oracles $\mathcal{H}_x^{(k+1)}$ by a Monte Carlo approximation of the expectation $\bar{\mathsf{S}}_x(\hat{\Lambda}^{(k)})$ (see (7))

$$\mathcal{H}_x^{(k+1)} := \frac{1}{M}\sum_{m=1}^{M} S_x(\mathbf{R}^{m,k+1}, \mathbf{O}^{m,k+1}, \mathbf{B}^{m,k+1}),$$

where $\{(\mathbf{R}^{m,k+1}, \mathbf{O}^{m,k+1}, \mathbf{B}^{m,k+1}), m\geqslant 0\}$ is the path of a Markov chain designed to have $\pi(\cdot|\mathsf{Z},\mathsf{I};\hat{\Lambda}^{(k)})$ as the unique invariant distribution. Here again, we use the `Gibbs-PGdual` algorithm (see Section II-B).

Convergence analysis of SAEM when the random oracles are *biased* approximations of the quantity of interest (note that they rely on a (non-stationary) Markov chain) is out of the scope of this paper; the interested reader can refer to [26, Section III-B-3] or [27].

III. Covid-19 monitoring

Let us apply the methodologies detailed in Section II to the estimation of the Covid19 pandemic reproduction number. The data are those made available at Johns Hopkins University repository; it collects every day since the earliest phase of the pandemic, and until mid-march 2023, the Covid-19 data made available by the National Health Authorities of more than 200 countries worldwide. The chronological series Z used in this work, are the daily new infections counts in France, for $T = 68$ days starting on April 29, 2022. They are displayed in Fig. 1[right] (black curve); the counts look under-estimated one or two days a week, over-estimated few days after, and the errors look small between successive periods of such a strong variability. This calls for testing the mixture model on the a priori distribution of the errors, O_t, introduced in Section II-A.

Parameter estimation by SAEM. SAEM is run with a step size $\gamma_x^{(k)}$ which is constant during the first iterations and then decreasing at the rate $1/\sqrt{k}$ until $k_{\max} = 15000$ iterations. The oracles $\mathcal{H}_x^{(k)}$ are computed with $M = 375\,000$ samples. A path of SAEM, $k\mapsto\hat{\Lambda}^{(k)} := (\hat{\lambda}_{\mathsf{R}}^{(k)}, \hat{\lambda}_{\mathsf{O},1}^{(k)}, \hat{\lambda}_{\mathsf{O},2}^{(k)}, \hat{\omega}^{(k)})$ is displayed in red in Fig. 1[left]. The SAEM sequence $\{\hat{\Lambda}^{(k)}, k\geq 0\}$ converges towards the limiting value $\widehat{\Lambda} := (573, 9.60\,10^{-5}, 3.28\,10^{-5}, 0.26)$. Observe that $\hat{\lambda}_{\mathsf{O},1}/\hat{\lambda}_{\mathsf{O},2}\approx 3$ so that, under the a priori distribution, the expectation and the

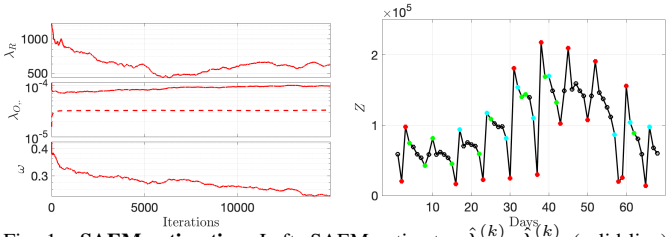

Fig. 1. **SAEM estimation.** Left: SAEM estimates $\hat{\lambda}_{\mathsf{R}}^{(k)}$, $\hat{\lambda}_{\mathsf{O},1}^{(k)}$ (solid line), $\hat{\lambda}_{\mathsf{O},2}^{(k)}$ (dashed line) and $\hat{\omega}^{(k)}$, as functions of the number of iterations k. Right: Counts Z_t marked in red, cyan, green and black, when the a posteriori probability that $|\mathsf{O}_t|$ is large (i.e. $\mathsf{B}_t = 0$), is respectively in $[0.9, 1]$, in $[0.8, 0.9]$, in $[0.7, 0.8]$, and less than 0.7.

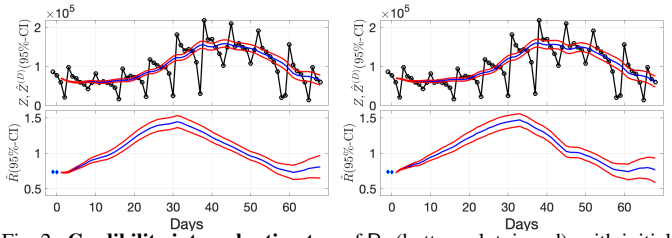

Fig. 2. **Credibility interval estimates:** of R_t (bottom plot, in red), with initial values $\mathsf{R}_0, \mathsf{R}_{-1}$ shown as blue diamonds; and of denoised counts $\mathsf{Z}_t^{(D)}$ (top plot, in red), superimposed to counts Z_t (black line). In blue, the estimate of the median. Two models are compared: the mixture model for the a priori distribution of O_t (left) and the model in [12] (right).

standard deviation of the absolute value of the errors, $|\mathsf{O}_t|$, are larger in class $\#2$ than in class $\#1$, by a factor 3.

Given the estimation $\widehat{\Lambda}$ of the HMM parameters, we use the a posteriori distribution $\pi(\mathbf{R}, \mathbf{O}, \mathbf{B}|\mathbf{Z}, \mathsf{I}; \widehat{\Lambda})$ to obtain information on the hidden variables $(\mathbf{R}, \mathbf{O}, \mathbf{B})$. This distribution is approximated by the `Gibbs-PGdual` Markov chain $\{\mathbf{H}^{m,\infty} := (\mathbf{R}^{m,\infty}, \mathbf{O}^{m,\infty}, \mathbf{B}^{m,\infty}), m \geq 0\}$ (see Section II-B).

Classification. For each day $\#t$, we compute $M^{-1}\sum_{m=1}^{M} \mathbb{1}_{\mathsf{B}_t^{m,\infty}=0}$, a Monte Carlo approximation of $\mathbb{P}(\mathsf{B}_t = 0|\mathbf{Z}, \mathsf{I}; \widehat{\Lambda})$, the a posteriori probability that the count data Z_t is associated to a large absolute error $|\mathsf{O}_t|$ (i.e. from the component $\#2$ of the a priori mixture distribution). These probabilities are displayed in Fig. 1[right]. The plot illustrates that the HMM with a a priori mixture distribution for the errors, has an excellent comprehension of the data: the counts Z_t which look aberrant have a high probability to be associated to a large error.

Credibility intervals. The Markov chain $\{\mathbf{H}^{\mathbf{m},\infty}, m = 0, \cdots, 10^7\}$ is now used to estimate the quantiles $q_\alpha(\mathsf{R}_t)$ and $q_\alpha(\mathsf{O}_t)$ of order $\alpha \in \{0.025, 0.5, 0.975\}$ of the a posteriori distribution of R_t and O_t. The estimators are the empirical quantile ones. We deduce a CI at the asymptotic level 0.95 for R_t and for O_t from the estimations of the quantiles $q_{0.025}(\cdot)$ and $q_{0.975}(\cdot)$; and a point estimate from the estimation of $q_{0.5}(\cdot)$. However, these empirical quantiles depend on the randomness induced by the Markov chain $\{\mathbf{H}^{\mathbf{m},\infty}, m \geq 0\}$; the mean value $\mu[q_\alpha(L_t)]$ is estimated, for $L_t \in \{\mathsf{R}_t, \mathsf{O}_t\}$, from 200 empirical quantiles computed from 200 independent realizations of the Markov chain $\{\mathbf{H}^{\mathbf{m},\infty}, m \geq 0\}$. In Fig. 2[left, bottom], the mean value $\mu[q_{0.5}(\mathsf{R}_t)]$ of the Monte Carlo estimate of the median $q_{0.5}(\mathsf{R}_t)$ is displayed in blue, together with the mean values $\mu[q_{0.025}(\mathsf{R}_t)]$ and $\mu[q_{0.975}(\mathsf{R}_t)]$ (in red). The same analysis is performed for the errors O_t, from which we deduce mean values of the estimates of the quantiles for the *denoised data* $\mathsf{Z}_t^{(D)} := \mathsf{Z}_t - \mathsf{O}_t$ (see Fig. 2[left,top]). The first conclusion is that, when t is small, the CIs for R_t are influenced by the initial values R_ℓ, $\ell \in \{-1, 0\}$, fixed here to $\mathsf{Z}_\ell/\Phi_\ell^{\mathbf{Z}}$ (see [10, Section 2.2.]): $\mathsf{R}_0 = 0.73$ and $\mathsf{R}_{-1} = 0.74$. Second, in adequacy with the evolution of the counts Z_t, the reproduction number increases and then decreases: note that this change of monotonicity for $t \mapsto \mathsf{R}_t$ occurs about ten days before the counts data express this change.

Benefits of the mixture model. The mixture model detailed in Section II-A extends the simpler model in [12] where the a priori distribution on O_t is a Laplace distribution. We calibrate this simpler model by running a SAEM algorithm (detailed derivations of the algorithm and numerical convergence of the SAEM sequence are not shown here). We then repeat the computation of the quantiles as explained above for the mixture model: the mean value of the quantiles, computed from 200 independent realizations, is displayed in Fig. 2[right] and yields the following conclusions. First, the two models produce consistent estimates for $t \mapsto \mathsf{R}_t$: the estimates of R_t increase until day $\#t = 30$, reach a maximal value around day $\#t = 32$, and then decrease. Second, CI for R_t obtained by the two models identifies similar ranges; yet the mixture model succeeds better in capturing uncertainties on the values of the errors: a close inspection indeed shows that whatever the time t, the mixture model yields narrower CIs than the ones from simpler model, the ratio of the sizes being lower than 0.94 after day $\#t = 4$ and even about 0.8 around day $\#t = 50$. Finally, contrary to the simpler model, the mixture model allows to quantify the probability that a count Z_t is corrupted by a large error. However, all these benefits have to be put next to the computational cost, which increases because of the larger number of parameters (four instead of two), and of the introduction of T additional hidden variables $\mathbf{B}$ (the Markov chain takes values in $\mathbb{R}^{2T} \times \{0, 1\}^T$ instead of $\mathbb{R}^{2T}$) which increases the Monte Carlo cost per iteration of the MCMC sampler.

IV. Conclusion

These promising preliminary achievements will be complemented by exploring several tracks. First, sequential statistical modeling and learning for an online processing of the daily counts will be considered. Second, we will explore alternatives strategies for data-driven automated parameter selection. Adopting a hierarchical Bayes formalism [28, Ch 10], the proposed HMM will be extended by also considering Λ as a random variable which can be estimated by MCMC algorithms. Another strategy is to assume the existence of a deterministic ground truth for R_t, O_t, and to augment the variational estimation procedure of [10] with an adapted Stein Unbiased Risk Estimate [29], [30] enabling automated data-driven selection of regularization parameters.

REFERENCES

[1] A. Flahault, "COVID-19 cacophony: is there any orchestra conductor?," *The Lancet*, vol. 395, no. 10229, pp. 1037, 2020.

[2] J. Arino, "Describing, modelling and forecasting the spatial and temporal spread of COVID-19–A short review," Tech. Rep., arXiv:2102.02457, 2021.

[3] Q.-H. Liu, M. Ajelli, A. Aleta, S. Merler, Y. Moreno, and A. Vespignani, "Measurability of the epidemic reproduction number in data-driven contact networks," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 115, pp. 12680–12685, 2018.

[4] F. Brauer, C. Castillo-Chavez, and Z. Feng, *Mathematical models in epidemiology*, Springer, New York, 2019.

[5] O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz, "On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations," *J. Math. Biol.*, vol. 28, pp. 365–382, 1990.

[6] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Math Biosci*, vol. 180, pp. 29–48, 2002.

[7] T. Obadia, R. Haneef, and P.-Y. Boëlle, "The R0 package: A toolbox to estimate reproduction numbers for epidemic outbreaks," *BMC Medical Inform Decis. Mak.*, vol. 12, pp. 147, 2012.

[8] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, "A new framework and software to estimate time-varying reproduction numbers during epidemics," *Am. J. Epidemiol.*, vol. 178, pp. 1505–1512, 2013.

[9] P. Abry et al., "Spatial and temporal regularization to estimate COVID-19 reproduction number $R(t)$: Promoting piecewise smoothness via convex optimization," *PLOS One*, vol. 15, 2020, e0237901.

[10] B. Pascal, P. Abry, N. Pustelnik, S. Roux, R. Gribonval, and P. Flandrin, "Nonsmooth convex optimization to estimate the covid-19 reproduction number space-time evolution with robustness against low quality data," *IEEE Trans. Signal Process.*, vol. 70, pp. 2859–2868, 2022.

[11] O. Cappé, E. Moulines, and T. Ryden, *Inference in Hidden Markov Models*, Springer Series in Statistics. Springer, 2005.

[12] G. Fort, B. Pascal, P. Abry, and N. Pustelnik, "Covid19 Reproduction Number: Credibility Intervals by Blockwise Proximal Monte Carlo Samplers," *IEEE Trans. Signal Process.*, vol. 71, pp. 888–900, 2023.

[13] P. Abry, G. Fort, B. Pascal, and N. Pustelnik, "Credibility intervals for the reproduction number of the Covid-19 pandemic using Proximal Langevin samplers," in *2023 31th European Signal Processing Conference (EUSIPCO)*, 2023, Accepted.

[14] H. Robbins and S. Monro, "A Stochastic Approximation method," *Ann. Math. Stat.*, pp. 400–407, 1951.

[15] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Roy. Stat. Soc. B Met.*, vol. 39, no. 1, pp. 1–38, 1977.

[16] P. Abry, G. Fort, B. Pascal, and N. Pustelnik, "Temporal evolution of the Covid19 pandemic reproduction number: Estimations from proximal optimization to Monte Carlo sampling," in *Annu. Int. Conf. IEEE Eng. Med. Biol.-Proc.*, 2022, pp. 167–170.

[17] H. Artigas, B. Pascal, G. Fort, P. Abry, and N. Pustelnik, "Credibility interval Design for Covid19 Reproduction Number from Nonsmooth Langevin-type Monte Carlo sampling," in *2022 30th European Signal Processing Conference (EUSIPCO)*, 2022, pp. 2196–2200.

[18] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive algorithms and stochastic approximations*, vol. 22, Springer Science & Business Media, 2012.

[19] T.L. Lai, "Stochastic approximation," *Ann. Stat.*, vol. 31, no. 2, pp. 391–406, 2003.

[20] B. Delyon, M. Lavielle, and E. Moulines, "Convergence of a stochastic approximation version of the EM algorithm," *Ann. Stat.*, pp. 94–128, 1999.

[21] J.-Y. Tourneret, M. Doisy, and M. Lavielle, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise: application to SAR image edge detection," *Signal Process.*, vol. 83, no. 9, pp. 1871–1887, 2003.

[22] F. Septier, Y. Delignon, A. Menhaj-Rivenq, and C. Garnier, "Monte Carlo Methods for Channel, Phase Noise, and Frequency Offset Estimation With Unknown Noise Variances in OFDM Systems," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3613–3626, 2008.

[23] A. Boisbunon and J. Zerubia, "Estimation of the weight parameter with SAEM for marked point processes applied to object detection," in *2014 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 2185–2189.

[24] J. Liu, S. Kumar, and D. P. Palomar, "Parameter Estimation of Heavy-Tailed AR Model With Missing Data Via Stochastic EM," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 2159–2172, 2019.

[25] R. Zhou, J. Liu, S. Kumar, and D. P. Palomar, "Student's t var modeling with missing data via stochastic EM and Gibbs sampling," *IEEE Trans. Signal Process.*, vol. 68, pp. 6198–6211, 2020.

[26] A. Dieuleveut, G. Fort, E. Moulines, and H.-T. Wai, "Stochastic Approximation Beyond Gradient for Signal Processing and Machine Learning," *IEEE Trans. Signal Process.*, vol. 71, pp. 3117–3148, 2023.

[27] S. Allassonnière and J. Chevallier, "A new class of stochastic EM algorithms. Escaping local maxima and handling intractable sampling," *Comput. Stat. Data Anal.*, vol. 159, pp. 107159, 2021.

[28] C. P. Robert, *The Bayesian choice: from decision-theoretic foundations to computational implementation*, Springer, 2007.

[29] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, pp. 1135–1151, 1981.

[30] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, 2008.