STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–19.
Published online in International Academic Press (www.IAPress.org)

Bayesian and Likelihood Inference for the SIR Model Using Skellam's Distribution with Real Application to COVID-19

AbdeLati Lagzini [1,*], Hamid El Maroufy [1], Abdelakrim Merbouha [1], Mohamed El Omari [2]

[1] *Department of Applied Mathematics, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco*
[2] *Polydisciplinary Faculty of Sidi Bennour, Chouaib Doukkali University, Morocco*

Abstract In this paper, we focus on the well-known SIR epidemic model, formulated as a Markov counting process with the discrete Skellam distribution. Our main objective is to estimate its key parameters, namely the infection and recovery rates. We develop a Bayesian approach that relies on Markov chain Monte Carlo and data-augmentation techniques, and establish the posterior distributions under suitable priors. We then compare the Bayesian estimators with maximum likelihood (ML) estimators, for which we study weak consistency and asymptotic normality. Finally, the theoretical results are supported with numerical simulations and illustrated through a real-world application to COVID-19 data from Morocco.

Keywords SIR model, Bayesian inference, MCMC, MLE, Skellam distribution, missing data.

AMS 2010 subject classifications: 92D30, 62F15, 60J10, 62M10.

DOI: 10.19139/soic-2310-5070-2993

1. Introduction

Advances in mathematical and computational approaches in recent years have made possible a more rigorous analysis of epidemics. Consequently, different approaches have led to the proposal of numerous epidemic models, including the widely studied **SIR** (susceptible–infected–recovered) model [19]. As described by Bailey in [7], this model is utilized to analyze the progression of complex infectious diseases and assess the effects of public health programs and interventions.

There are several types of parameter estimation approach, such as the Bayesian approach [26] and the maximum likelihood (ML) approach [11]. The ML method is consistent and relies on realizations x_k derived from observations X_k, where $k = 0, 1, \ldots, N$. Moreover, the Bayesian approach can be used: it consists of incorporating additional information about the population's parameters through prior knowledge, which can be inferred from preliminary study [9]. This approach is known as a powerful estimation method that involves tedious computational techniques, including the MCMC (Markov Chain Monte Carlo) techniques [25].

In this work, we use the aforesaid estimation method to infer the unknown infection and removed rates in the stochastic epidemic model. We do this by characterizing the state process, by Skellam distribution [30]. This distribution is defined as the law of the difference of two independent random variables [4]. This distribution is particularly relevant for modeling scenarios in which positive and negative events interact significantly, as in the propagation and recovery processes of an epidemic like the **SIR** model.

*Correspondence to: Abdelati Lagzini (Email: abdelati.lagzini@usms.ma). Department of Applied Mathematics, Faculty of Science and Technology, Sultan Moulay Slimane University, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

The approaches described in this paper are distinguished by their exact and rigorous formulation of the likelihood function, given by the Skellam distribution. This fact sets our work apart from studies, that are based on approximation likelihood, as in [24]. Due to their flexibility and conjugacy, we use Gamma distributions as priors of the infection rate β and recovery rate γ [21], this choice is well-suited for positive parameters and simplifies the computational aspect of the posteriors. Moreover, to improve the accuracy of these Bayesian estimators, when the observation times are sparse, we follow the technique built in our previous work on the SIS epidemic model [12]; we introduce $M-1$ of latent values points between each consecutive observation, which contributes to a better understanding of the unobserved events arising during the epidemic. For further details, the readers should view the studies in [13, 21]. To illustrate the efficiency of the Bayesian estimator compared to the obtained ML estimator, for which, we prove the weak consistency and asymptotic normality; numerical simulations are made. As well as we apply the proposed methods to real data, that concerns the daily number of infected and recovery individuals of the COVID-19 pandemic in Morocco.

We organize the rest of the article as follows: In the next section, we introduce the Skellam distribution as a bivariate distribution. In Section 3, we describe the dynamics of the SIR epidemic model using the Skellam distribution. We explore the Bayesian inference as an approach for model parameter estimation in Section 4. In Section 5, we study the asymptotic proprieties of the MLE. In Section 6, we support the theoretical results with numerical simulations and present a real-world application. Finally, the paper terminates with a conclusion and appendix sections.

Throughout the paper, we denote $\mathcal{P}ois(\alpha)$ as the Poisson distribution with parameter α, and $\mathbb{P}(.)$ and $\pi(.)$ represent, respectively, the normalized and non-normalized distributions.

2. The Bivariate Skellam Distribution

We first define the univariate Skellam distribution before introducing the bivariate case.

Definition 1
Let W_1 and W_2 be two independent Poisson random variables, such that $W_1 \sim \mathcal{P}ois(\lambda_1)$ and $W_2 \sim \mathcal{P}ois(\lambda_2)$. The difference $X = W_1 - W_2$ is distributed as Skellam distribution, denoted $X \sim \mathcal{S}(\lambda_1, \lambda_2)$, with probability mass function is given by

$$\mathbb{P}(X = x \mid \lambda_1, \lambda_2) = \exp\left(-\lambda_1 - \lambda_2\right) \left(\frac{\lambda_1}{\lambda_2}\right)^{x/2} \mathfrak{I}_{|x|}(2\sqrt{\lambda_1\lambda_2}), \quad x \in \mathbb{Z}, \tag{1}$$

where $\mathfrak{I}_y(z)$, present the modified Bessel function of the first kind, is defined as $\mathfrak{I}_y(z) = \sum_{p=0}^{\infty} \frac{1}{p!\Gamma(y+p+1)} \left(\frac{z^2}{4}\right)^{p+y}$.

Early in 1937, Irwin [15] developed the distribution of the difference between two iid Poison random variables ($\lambda_1 = \lambda_2$). The situation where $\lambda_1 \neq \lambda_2$ was studied by Skellam [30] and Prekopa [22] (the reader can consult [16] for more details).

For the bivariate case, let Y_i, $i \in \{0, 1, 2\}$, be three independent Poisson-random variables with respective parameters λ_0, λ_1, and λ_2. We define

$$X_1 = Y_1 + \delta_1 Y_2, \quad X_2 = Y_0 + \delta_2 Y_2, \tag{2}$$

where δ_1 and δ_2 are constants taking only the values -1 or $+1$.

In this work, we are particularly interested in the case where $Y_0 \equiv 0$, and $(\delta_1, \delta_2) = (-1, +1)$.

So (2), becomes

$$X_1 = Y_1 - Y_2, \; X_2 = Y_2, \tag{3}$$

a simple calculations means that, for $(k,r) \in \mathbb{Z} \times \mathbb{N}$ and $r \geq \min(0,-k)$

$$\mathbb{P}(X_1 = k, X_2 = r) = \exp\left(-\lambda_1 - \lambda_2\right) \frac{\lambda_1^{k+r}}{(k+r)!} \frac{\lambda_2^r}{r!}. \tag{4}$$

3. Characterization of the SIR model by Poisson process

Within the standard **SIR** epidemic model, individuals are classified into three groups: susceptible, infective, and recovered. The population under consideration is assumed to be closed and homogeneous, with a total size of $n+a$. At any time $t \geq 0$, the corresponding group sizes are denoted by $S(t)$, $I(t)$, and $R(t)$, with the constraint $I(t) + S(t) + R(t) = n + a$.

At the initial time, only susceptibles and infectives are present, namely $(I(0), S(0), R(0)) = (a, n, 0)$. The epidemic process is described by $\mathbf{X}_n(t) = \{(I(t), R(t)); t \geq 0\}$, which is formulated as a two-dimensional continuous-time Markov chain with state space $\mathbb{S} = \{(i,r)\,;\, 0 \leq i \leq n,\, 0 \leq r \leq (n-i)+a\}$. We then have the following transition probabilities for a small δt:

$$\begin{aligned}
&\mathbb{P}\Big(\mathbf{X}_n(t+\delta t) = (i+1, r) \mid \mathbf{X}_n(t) = (i,r)\Big) = \frac{\beta i}{n}\big(a + n - r - i\big)\delta t + o(\delta t),\\
&\mathbb{P}\Big(\mathbf{X}_n(t+\delta t) = (i-1, r+1) \mid \mathbf{X}_n(t) = (i,r)\Big) = \gamma i \delta t + o(\delta t),\\
&\mathbb{P}\Big(\mathbf{X}_n(t+\delta t) = (i, r) \mid \mathbf{X}_n(t) = (i,r)\Big) = 1 - \left(\frac{\beta i\big(a+n-r-i\big)}{n} + \gamma i\right)\delta t + o(\delta t).
\end{aligned} \tag{5}$$

The probability of all other transitions is $o(\delta t)$. The parameters β and γ present, respectively, the infection and recovery rates. The process is ending whenever the number of infected individuals reaches zero, see [9, 24] and the references therein.

The original Markov process characterized by (5), can be formulated as sum of independent Poisson process; according to [5, p. 40], see also [17, Chapter 11], $\mathbf{X}_n$ can be rewritten

$$\mathbf{X}_n(t) = \mathbf{X}_n(0) + \sum_{\ell} Y_\ell \left(n \int_0^t B_\ell \left(n^{-1} \mathbf{X}_n(s) \right) ds \right) \ell, \tag{6}$$

where the non-random starting point is $\mathbf{X}_n(0) = (a, 0)$, and $Y_\ell := (Y_\ell(t); t \geq 0)$ are independent standard Poisson processes.

Note that, as described by Andersson and Briton in [5], the process is considered to allow a restricted set of transitions, represented by a finite number of vectors $\ell \in \mathbb{Z}^d$ such that $\sup_x B_\ell(x) > 0$; in particular, $\ell \in \{(-1,1),(1,0)\}$. The time-density function $B_\ell(.)$ is supported to be a continuous function. Therefore, the process $\mathbf{X}_n$ becomes

$$\begin{aligned}
\mathbf{X}_n(t) &= \mathbf{X}_n(0) + \sum_{\ell} \ell Y_\ell \left(\int_0^t n B_\ell \left(n^{-1} \mathbf{X}_n(s) \right) ds \right)\\
&= \mathbf{X}_n(0) + (-1,1)\, Y_{(-1,1)} \left(\frac{\beta}{n} \int_0^t I(s)\big(n + a - I(s) - R(s)\big) ds \right) + (1,0)\, Y_{(1,0)} \left(\gamma \int_0^t I(s) ds \right)\\
&= \begin{pmatrix} a + Y_{(1,0)} \left(n \int_0^t B_{(1,0)} \left(n^{-1} \mathbf{X}_n(s) \right) ds \right) - Y_{(-1,1)} \left(n \int_0^t B_{(-1,1)} \left(n^{-1} \mathbf{X}_n(s) \right) ds \right) \\ Y_{(-1,1)} \left(n \int_0^t B_{(-1,1)} \left(n^{-1} \mathbf{X}_n(s) \right) ds \right) \end{pmatrix}^\top\\
&= \begin{pmatrix} a + \mathcal{P}ois(\lambda_1) - \mathcal{P}ois(\lambda_2) \\ \mathcal{P}ois(\lambda_2) \end{pmatrix}^\top,
\end{aligned}$$

where the Poisson process $\mathcal{P}ois(\cdot)$ has rates $\lambda_1 = \frac{\beta}{n}\int_0^t I(s)\big(n+a-I(s)-R(s)\big)ds$ and $\lambda_2 = \gamma\int_0^t I(s)ds$, respectively, which is equivalent to (3). Therefore, we have:

$$\mathbb{P}\Big(\mathbf{X}_n(t) = (k,r) \mid \lambda_1, \lambda_2\Big) = \exp\left(-\lambda_1-\lambda_2\right)\frac{\lambda_1^{k+r-a}}{(k+r-a)!}\frac{\lambda_2^r}{r!}, \quad (k,r)\in\mathbb{S}. \tag{7}$$

In practice, a discrete version of (7) is essential because the process $(\mathbf{X}_n(t), t\geq 0)$ is sampled at discrete time points in the interval $[0,T]$, specifically at a given time $t_0 = 0 < t_1 < t_2 < \cdots < T = t_N$, for $t_i - t_{i-1} =: \delta t_i = \frac{T}{N}$ with a constant integer $N > 0$. This discretized form is provided by Euler's as detailed in [18, Algorithm 1]. For $i \in \{1,2,\ldots,N\}$, we obtain the expression:

$$\mathbb{P}\left(\mathbf{X}_n(t_i) = (k_i,r_i) \mid \mathbf{X}_n(t_{i-1}), \lambda_1^i, \lambda_2^i\right) = \exp\left(-\lambda_1^i-\lambda_2^i\right)\frac{(\lambda_1^i)^{k_i+r_i-a}}{(k_i+r_i-a)!}\frac{(\lambda_2^i)^{r_i}}{r_i!}, \tag{8}$$

where $(k_i,r_i)\in\mathbb{S}$, and the time rates λ_1^i and λ_2^i are given by the expressions

$$\lambda_1^i = \frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)\big(a+n-R(s)-I(s)\big)ds, \quad \lambda_2^i = \gamma\int_{t_{i-1}}^{t_i} I(s)ds.$$

By substituting these time rates into equation (8), it becomes:

$$\begin{aligned}\mathbb{P}\Big(\mathbf{X}_n(t_i) = (k_i,r_i) \mid \mathbf{X}_n(t_{i-1}), \beta, \gamma\Big) \quad = \quad & \exp\left(-\frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)\big(a+n-I(s)-R(s)\big)ds - \gamma\int_{t_{i-1}}^{t_i} I(s)ds\right) \\ & \frac{\left(\frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)\big(a+n-R(s)-I(s)\big)ds\right)^{k_i+r_i-a}}{(k_i+r_i-a)!} \\ & \frac{\left(\gamma\int_{t_{i-1}}^{t_i} I(s)ds\right)^{r_i}}{r_i!}.\end{aligned} \tag{9}$$

In the context of infectious disease modeling, the basic reproduction number R_0 plays a fundamental role and is expressed as $R_0 = \frac{\beta}{\gamma}$. This number determines whether an epidemic can occur by measuring the disease's transmission potential and evaluating its threshold behavior. The parameters β and γ must be estimated because they are unknown. This paper aims to demonstrate efficient methods that use ML and Bayesian inferences to estimate them, e.g., [3]. For the remainder of this document, we set $\theta = (\beta,\gamma)$.

4. Bayesian Framework

Although the model process (6) is defined in continuous-time, the data available in practice are generally obtained at discrete observation times, and often at relatively spaced frequencies. This creates a difficulty in estimating and inferring the parameters since it is assumed that the process $\mathbf{X}_n(t) = \{(I(t),R(t)), t\geq 0\}$ will be observed only a finite number of times. To overcome this difficulty, we augment the data using Eraker's approach [13]. This approach involves inserting $M-1$ latent points of data between consecutive observations, where $M\in\mathbb{N}^*$. It relies on MCMC methods and is applicable to various models, particularly those with missing data, see [24, 1].

Hence, to keep the discretization bias randomly small,the time step is defined as $\delta t = \frac{t_i - t_{i-1}}{M}$, where M is a strategically selected positive integer. The interval $[0,T]$ is then divided into $N = MT$ evenly spaced time points: $t_0 = 0 < \cdots < t_{M-1} < t_M < t_{M+1} < \cdots < T = t_N$

For this, we introduce the matrix $\hat{\mathbf{X}}$, which contains all components of the augmented dataset, that is, both the observed and the missing values; therefore, ...

$$\hat{\mathbf{X}} = \begin{pmatrix} I_{t_0} & \hat{I}_{t_1} & \cdots & \hat{I}_{t_{M-1}} & I_{t_M} & \hat{I}_{t_{M+1}} & \cdots & I_{t_N} \\ R_{t_0} & \hat{R}_{t_1} & \cdots & \hat{R}_{t_{M-1}} & R_{t_M} & \hat{R}_{t_{M+1}} & \cdots & R_{t_N} \end{pmatrix}.$$

If i is an integer multiple of M, then $\hat{X}_i$ represents an observed data value, equivalently written as $\hat{\mathbf{X}}^{\text{obs}}$. On the other hand, if i is not a multiple of M, $\hat{X}_i$ corresponds to the missing data (no-obs), are denoted by $\hat{\mathbf{X}}^{\text{no-obs}}$. Typically, the notation $\hat{X}_i$ can refer to either missing or observed data, depending on the context. In total, there are $2(M-1)T$ data points that remain unobserved in the system. Thus, conditional on the first observation, the joint posterior density can be written as

$$\pi(\hat{\mathbf{X}}, \theta) \propto \prod_{i=1}^{N} \mathbb{P}\left(\hat{X}_i \mid \hat{X}_{i-1}, \beta, \gamma\right) \pi(\theta), \tag{10}$$

where the distribution of $\left(\hat{X}_i \mid \hat{X}_{i-1}, \theta\right)$ is given in Eq.(9), and $\pi(\theta)$ is the joint prior distribution for β and γ.

In Eq. (10), we establish the joint posterior law of the model parameters based on the combination of observed and latent data. However, we are mainly interested in the distribution of $\left(\theta, \hat{\mathbf{X}}^{\text{no-obs}} \mid \hat{\mathbf{X}}^{\text{obs}}\right)$. As described in [9, 10] and references therein, inference can be performed by alternating simulations of unobserved data and parameter estimates given the augmented data and the current parameter values.

Given the large amount of missing data in epidemiological models, a Gibbs sampler is particularly effective for sampling variables. At the first step, a candidate value for (β, γ) is sampled according to the joint prior distribution $\pi(\theta)$, which is then employed to generate the latent trajectory, as explained in the subsequent section. We then update $\theta = (\beta, \gamma)$ with a newly generated value and adjust the latent observations accordingly. By successive repetition of this update scheme, we obtain a Markov chain that converges in distribution to the posterior $\pi\left(\theta, \hat{\mathbf{X}}^{\text{no-obs}} \mid \hat{\mathbf{X}}^{\text{obs}}\right)$, which constitutes its unique stationary distribution [32].

4.1. The posterior distributions of missing data

The initial stage in the Gibbs sampler is to update the missing data paths by using Eraker's approach [13] to generate one column of latent observations $\hat{X}_i$ given $\hat{X}_{i-1}$ and $\hat{X}_{i+1}$, based on the conditional distribution

$$\pi\left(\hat{X}_i \mid \hat{X}_{\setminus i}, \theta\right) \propto \pi\left(\hat{X}_i \mid \hat{X}_{i-1}, \hat{X}_{i+1}, \theta\right), \tag{11}$$

where $\hat{X}_{\setminus i}$ denotes all columns of $\hat{\mathbf{X}}$ except the i-th one. The explicit form of this conditional density, proved in Appendix A, is stated as follows:

Theorem 1
When δt is sufficiently small, we have

$$\begin{aligned}\pi\Big(\hat{X}_i = (x, y) \mid \hat{X}_{i-1}, \hat{X}_{i+1}, \theta\Big) \quad &\propto \beta^{x+y}\gamma^{y}\left(\left[\frac{\delta t}{2n}x(n+a-x-y)\right] + C_1^i\right)^{x+y-a} \\ &\times \left(\left[\frac{\delta t}{2n}x(n+a-x-y)\right] + C_2^i\right)^{C_5^i} \\ &\times \frac{\left(\frac{\delta t}{2}x + C_3^i\right)^{y}\left(\frac{\delta t}{2}x + C_4^i\right)^{R_{t_{i+1}}}}{(x+y-a)!y!},\end{aligned} \tag{12}$$

where C_j^i, $j \in \{1, \ldots, 5\}$, are constants calculated from the data, and its form are given by (25).

It is not possible to sample directly from this posterior distribution of latent data, so we employ an Accept–Reject Metropolis–Hastings (AR–MH) update at this stage [25, Chap. 6]. In this scheme, a new candidate $X_i' = (x', y')$ for $\hat{X}_i = (I_{t_i}, R_{t_i})$ is proposed from a discrete uniform distribution centered on the mean of its neighboring states $\hat{X}_{i-1}$ and $\hat{X}_{i+1}$, within the feasible SIR domain $\mathbb{S}$ defined previously in Section 3.

Denoting $\mu_I = \lfloor (I_{t_{i-1}} + I_{t_{i+1}})/2 \rfloor$ and $\mu_R = \lfloor (R_{t_{i-1}} + R_{t_{i+1}})/2 \rfloor$, we define the adaptive proposal windows by

$$w_I = \max\left(1,\ \text{round}(\eta\,|I_{t_{i+1}} - I_{t_{i-1}}|)\right), \quad w_R = \max\left(1,\ \text{round}(\eta\,|R_{t_{i+1}} - R_{t_{i-1}}|)\right), \quad \eta = 0.4,$$

and the proposal distribution is

$$X_i' \sim \text{Unif}\big(([\mu_I - w_I, \mu_I + w_I] \times [\mu_R - w_R, \mu_R + w_R]) \cap \mathbb{S}\big).$$

This proposal is symmetric by construction, leading to the Metropolis–Hastings acceptance probability

$$\alpha\left(X_i'; \hat{X}_i\right) = \min\left\{1, \frac{\pi\left(X_i' \mid \hat{X}_{i-1}, \hat{X}_{i+1}, \theta\right)}{\pi\left(\hat{X}_i \mid \hat{X}_{i-1}, \hat{X}_{i+1}, \theta\right)}\right\}.$$

At each sweep, the proposal widths (w_I, w_R) automatically expand when the local trajectory exhibits strong variability, and shrink when it stabilizes. This adaptive mechanism maintains acceptance probabilities within a desirable range. In our numerical experiments (Section 6), the empirical acceptance rates of the AR–MH update were 37% for $M = 2$, 45% for $M = 5$, and 52% for $M = 10$, which confirms that the sampler is well tuned and exhibits good mixing across the different interpolation levels. The complete implementation of the Gibbs sampler, including the adaptive AR–MH update for the latent states described in Subsection 4.1, is summarized in Algorithm 1.

4.2. Prior and posterior distributions of the parameters β and γ

In the final stage of the Gibbs sampler, we draw $\theta^{(h)} = (\beta^{(h)}, \gamma^{(h)})$, the parameter vector θ conditioned on its current value and the augmented data, at iteration h. Here, the parameters of interest must be positive, which leads us to choose a prior distribution that produces positive results. For this, we assume, as before, two independent Gamma distributions, *i.e.*, $\beta \sim \Gamma(m, \lambda)$ and $\gamma \sim \Gamma(m', \lambda')$. Due to conjugacy, this choice is practical for Bayesian inference. Moreover, the gamma distribution proves to be highly flexible and is commonly applied to describe the rate parameters of epidemic models. For further details, readers may consult [24] and the references therein. Bayes theorem leads to a gamma-type posterior distribution; the following theorem, whose full proof is presented in Appendix A, asserts this statement

Theorem 2

Suppose that β and γ are modeled with independent Gamma priors, namely $\beta \sim \Gamma(m, \lambda)$ and $\gamma \sim \Gamma(m', \lambda')$; then the posterior distributions of these parameters are given by

$$\beta \mid \hat{\mathbf{X}} \sim \Gamma\left(A_{3,N} + m - 1, A_{1,N} + \lambda\right), \tag{13}$$

$$\gamma \mid \hat{\mathbf{X}} \sim \Gamma\left(A_{4,N} + m' - 1, A_{2,N} + \lambda'\right), \tag{14}$$

where the constants $A_{1,N}, \cdots, A_{4,N}$ are given by

$$\left.\begin{aligned} A_{1,N} &= \int_0^T \frac{1}{n} I(t) S(t)\, \mathrm{d}t = \int_0^T \bar{I}(t)\big(a + n - R(t) - I(t)\big)\, \mathrm{d}t; \\ A_{2,N} &= \int_0^T I(t)\, \mathrm{d}t; \\ A_{3,N} &= \sum_{i=1}^{N} \left(I(t_i) + R(t_i) - a\right); \\ A_{4,N} &= \sum_{i=1}^{N} R(t_i). \end{aligned}\right\} \tag{15}$$

5. The maximum likelihood estimation for β and γ

In this section, we investigate the MLE [23] method; we aim to derive the most probable value for β and γ using a discrete observations of the process obtained above $\mathbf{X}_n(t) = ((I(t), R(t)), t \geq 0)$ at time points $0 =$

$t_0, t_1, \ldots, t_N = T$. Given the distribution specified in (9), the likelihood function of the vector of observation $\mathbf{X}_n = (X_n(t_0), X_n(t_1), \ldots, X_n(t_N))$ is then given as follow

$$\begin{aligned} L(\mathbf{X}_n \mid \theta) &= \prod_{i=1}^{N} \mathbb{P}\big(X_n(t_i) = (I(t_i), R(t_i)) \mid X_n(t_{i-1}), \theta\big) \\ &= \prod_{i=1}^{N} \exp\left(-\beta \int_{t_{i-1}}^{t_i} \frac{I(s)}{n}\,(a+n-R(s)-I(s))\,\mathrm{d}s - \gamma \int_{t_{i-1}}^{t_i} I(s)\,\mathrm{d}s\right) \\ &\quad \frac{\left(\frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)\,(a+n-R(s)-I(s))\,\mathrm{d}s\right)^{I(t_i)+R(t_i)-a}}{(I(t_i)+R(t_i)-a)!} \\ &\quad \frac{\left(\gamma \int_{t_{i-1}}^{t_i} I(s)\,\mathrm{d}s\right)^{R(t_i)}}{R(t_i)!}. \end{aligned} \tag{16}$$

The log-likelihood function $\ell_N(\theta)$ is simplified as

$$\ell_N(\theta) = \log\left(L(\mathbf{X}_n \mid \theta)\right) = -\beta A_{1,N} + \log(\beta) A_{3,N} - \gamma A_{2,N} + \log(\gamma) A_{4,N} + A_{5,N}, \tag{17}$$

where the constants $A_{1,N}, \cdots, A_{4,N}$ are predefined previously in (15), and

$$\begin{aligned} A_{5,N} = \sum_{i=1}^{N} \left[(I(t_i)+R(t_i)-a)\log\left(\int_{t_{i-1}}^{t_i} \bar{I}(s)\big(a+n-I(s)-R(s)\big)\,\mathrm{d}s\right) - \log\left((I(t_i)+R(t_i)-a)!\right)\right] \\ - \sum_{i=1}^{N}\left[R(t_i)\log\left(\int_{t_{i-1}}^{t_i} I(s)\,\mathrm{d}s\right) - \log\left(R(t_i)!\right)\right]. \end{aligned}$$

The maximum likelihood estimators of β and γ are, respectively, given by

$$\left.\begin{aligned} \widehat{\beta}_{ML} &= \frac{A_{3,N}}{A_{1,N}} = \frac{\sum_{i=1}^{N}(I(t_i)+R(t_i)-a)}{\int_0^T \bar{I}(s)\big(a+n-R(s)-I(s)\big)\,\mathrm{d}s} \\ \widehat{\gamma}_{ML} &= \frac{A_{4,N}}{A_{2,N}} = \frac{\sum_{i=1}^{N} R(t_i)}{\int_0^T I(s)\,\mathrm{d}s}. \end{aligned}\right\} \tag{18}$$

5.1. Consistency and Asymptotic Normality

In this subsection, we examine the weak consistency and asymptotic Gaussian distribution of the maximum likelihood estimator $\widehat{\theta}_{ML} = (\widehat{\beta}_{ML}, \widehat{\gamma}_{ML})$, which are particularly useful for inference in stochastic processes. Adopting Crowder's notation [8], we define the following terms: $\theta_0 = (\beta_0, \gamma_0)$ denotes the true value of the parameter, $\ell'_N(\theta)$ represents the gradient of $\ell_N(\theta)$, and $\ell''_N(\theta)$ corresponds to the matrix of second-order derivatives of $\ell_N(\theta)$. The information matrix is given by $B_N = \mathbb{E}\left[-\ell''_N(\theta_0)\right]$. Finally, $\{c_N\}$ represents a real sequence that will be specified later, such that $c_N \to \infty$.

Proving the weak consistency of the estimator $\widehat{\theta}_N$ reduces to verifying [8, formula (2.3)]. Specifically, for some Δ and a sequence (c_N) (neither Δ nor c_N depending on θ), we must show that, when $\|\theta - \theta_0\| = \delta_1 \leq \Delta$, we have, for some $\tilde{\theta} = (\tilde{\beta}, \tilde{\gamma})$ in the segment line for each row:

$$\mathbb{P}\left(-c_N^{-1/2}(\theta-\theta_0)^\top B_N^{-1/2} \ell''_N(\tilde{\theta})(\theta-\theta_0) \geq \delta_1^2\right) \to 1, \quad \text{as } N \to \infty. \tag{19}$$

We begin our computations as follows:

$$B_N^{-1/2}\ell_N''(\tilde{\theta}) = -\begin{pmatrix} \frac{\mathbb{E}[A_{3,N}]}{\beta_0^2} & 0 \\ 0 & \frac{\mathbb{E}[A_{4,N}]}{\gamma_0^2} \end{pmatrix}^{-1/2} \begin{pmatrix} \frac{A_{3,N}}{\tilde{\beta}^2} & 0 \\ 0 & -\frac{A_{4,N}}{\tilde{\gamma}^2} \end{pmatrix}$$
$$= -\begin{pmatrix} \frac{\beta_0}{\tilde{\beta}^2}\frac{A_{3,N}}{\sqrt{\mathbb{E}[A_{3,N}]}} & 0 \\ 0 & \frac{\gamma_0}{\tilde{\gamma}^2}\frac{A_{4,N}}{\sqrt{\mathbb{E}[A_{4,N}]}} \end{pmatrix}.$$

Therefore,

$$-c_N^{-1/2}(\theta-\theta_0)^\top B_N^{-1/2}\ell_N''(\tilde{\theta})(\theta-\theta_0)$$
$$= c_N^{-1/2}\left(\frac{\beta_0}{\tilde{\beta}^2}\frac{A_{3,N}}{\sqrt{\mathbb{E}\left[A_{4,N}\right]}}(\beta-\beta_0)^2 + \frac{\gamma_0}{\tilde{\gamma}^2}\frac{A_{4,N}}{\sqrt{\mathbb{E}\left[A_{3,N}\right]}}(\gamma-\gamma_0)^2\right)$$
$$= \frac{\beta_0(\beta-\beta_0)^2}{\tilde{\beta}^2}\left(\frac{c_N^{-1/2}A_{3,N}}{\sqrt{\mathbb{E}\left[A_{4,N}\right]}}\right) + \frac{\gamma_0(\gamma-\gamma_0)^2}{\tilde{\gamma}^2}\left(\frac{c_N^{-1/2}A_{4,N}}{\sqrt{\mathbb{E}\left[A_{4,N}\right]}}\right).$$

By using ordinary calculus, we have almost surely, as $N \to \infty$.

$$\frac{T}{N}A_{3,N} \longrightarrow \int_0^T (I(t)+R(t)-a)\,dt, \text{ and } \frac{T}{N}A_{4,N} \longrightarrow \int_0^T R(t)\,dt. \tag{20}$$

Now, since $R(t)+I(t)+S(t) = n+a$, with $R(t), I(t), S(t)$ being nonnegative for all $t \in [0,T]$, we deduce that R and I are bounded. As a result, we have the uniform integrability of the sequences $\left\{\frac{T}{N}A_{3,N}\right\}$ and $\left\{\frac{T}{N}A_{4,N}\right\}$. Combining this with (20) we obtain [28, Theorem 5], when $N \to \infty$

$$\frac{T}{N}\mathbb{E}\left[A_{3,N}\right] \longrightarrow \int_0^T \mathbb{E}\left[I(t)+R(t)-a\right]\,dt, \text{ and } \frac{T}{N}\mathbb{E}\left[A_{4,N}\right] \longrightarrow \int_0^T \mathbb{E}\left[R(t)\right]. \tag{21}$$

It follows that, for N sufficiently large, we obtain the following approximation almost surely:

$$\frac{c_N^{-1/2}A_{3,N}}{\sqrt{\mathbb{E}\left[A_{3,N}\right]}} = \left(c_N^{-1/2}\sqrt{\frac{N}{T}}\right)\frac{\left(\frac{T}{N}A_{3,N}\right)}{\sqrt{\frac{T}{N}\mathbb{E}\left[A_{3,N}\right]}}$$
$$\sim \left(c_N^{-1/2}\sqrt{\frac{N}{T}}\right)\frac{\int_0^T (I(t)+R(t)-a)\,dt}{\int_0^T \mathbb{E}\left[I(t)+R(t)-a\right]dt}.$$

Similarly, we obtain $\frac{c_N^{-1/2}A_{4,N}}{\sqrt{\mathbb{E}[A_{4,N}]}} \sim \left(c_N^{-1/2}\sqrt{\frac{N}{T}}\right)\frac{\int_0^T R(t)dt}{\int_0^T \mathbb{E}\left[R(t)\right]dt}$. By choosing $c_N = \sqrt{N}$ we can assert that the sufficient condition (2.3) in [8] holds, the proof is complete, and we can now state the following theorem:

Theorem 3

The maximum likelihood estimator $\widehat{\theta}_{ML} = (\widehat{\beta}_{ML}, \widehat{\gamma}_{ML})$ is weakly consistent; that is,

$$\widehat{\theta}_{ML} \xrightarrow{\mathbb{P}} \theta_0, \quad \text{as } N \to \infty. \tag{22}$$

For the normality asymptotic behavior of the MLEs, we adopt Sweeting's notation [31]. We recall that $\ell_N''(\theta)$ is the Hessian matrix of the log-likelihood $\ell_N(\theta)$, and we have

$$-\ell_N''(\theta) = \begin{pmatrix} \frac{A_{3,N}}{\beta^2} & 0 \\ 0 & \frac{A_{4,N}}{\gamma^2} \end{pmatrix}, \quad \text{and} \quad A_N = \begin{pmatrix} \sqrt{\frac{N}{T}} & 0 \\ 0 & \sqrt{\frac{N}{T}} \end{pmatrix}.$$

Here, $A_{3,N}$ and $A_{4,N}$ are as defined in (15). We have A_N^{-1} converges to the zero matrix as $N \to \infty$. Furthermore,

$$W_N(\theta) := A_N^{-1}\left(-\ell_N''(\theta)\right)\left(A_N^{-1}\right)^\top = \begin{pmatrix} \frac{1}{\beta^2}\frac{T}{N}A_{3,N} & 0 \\ 0 & \frac{1}{\gamma^2}\frac{T}{N}A_{4,N} \end{pmatrix}$$
$$\xrightarrow{\text{a.s}} \begin{pmatrix} \frac{1}{\beta^2}\int_0^T \big(I(t)+R(t)-a\big)\,\mathrm{d}t & 0 \\ 0 & \frac{1}{\gamma^2}\int_0^T R(t)\,\mathrm{d}t \end{pmatrix} =: W(\theta), \quad \text{as } N \to \infty,$$

uniformly on every compact subset of the parameter space $\Theta = \{\theta = (\beta, \gamma) : \beta, \gamma > 0\}$. Because $t \mapsto \big(I(t) + R(t) - a\big)$ and $t \mapsto R(t)$ are strictly positive in $[0, T]$, the matrix $W(\theta)$ must be definite positive almost surely. Hence, the condition **C1** (Growth and convergence) in [31] holds. The continuity condition **C2** in [31] is trivial, since both $-\ell_N''(\theta)$ and A_N are uniformly continuous in (β, γ) in every compact set of $\Theta \subset \mathbb{R}^2$. Now, using [31, Theorems 1 and 2] combined with the continuous mapping theorem, we obtain:

$$\Big(W_N(\theta)^{1/2}A_N(\widehat{\theta}_{ML} - \theta),\ W_N(\theta)\Big) \xrightarrow{\mathcal{L}} (Z, W(\theta)), \quad \text{as } N \to \infty, \tag{23}$$

uniformly on every compact subset of Θ. Here, Z is a two-dimensional standard normal vector.

6. Simulation and Implementation

For the implementation of the estimation procedure described above, we propose the following algorithm, which provides a summary of the MCMC simulation approach based on the posteriors established in Theorems (1) and (2):

Algorithm 1 Gibbs Sampler with Adaptive AR-MH Update for Posterior Sampling

Initialization:
- Interpolate linearly between observed values of $\mathbf{X}_i$ to initialize latent trajectories $\widehat{\mathbf{X}}_i$.
- Set initial values for parameters $\beta^{(0)}$, $\gamma^{(0)}$, and choose the scaling factor $\eta = 0.4$.

for $h = 1, 2, \ldots, N_{\text{iter}} - 1$ **do**
  Step 1: Update latent data $\widehat{X}_i$ (AR–MH step).
  for $i = 2, \ldots, N-1$ **such that** $i \not\equiv 0 \pmod{M}$ **do**
    Propose a new latent state X_i' using the adaptive AR-MH rule defined in Subsection 4.1.
    Compute the acceptance probability $\alpha = \min\left\{1,\ \dfrac{\pi\left(X_i' | \hat{X}_{i-1}, \hat{X}_{i+1}, \theta^{(h)}\right)}{\pi\left(\hat{X}_i | \hat{X}_{i-1}, \hat{X}_{i+1}, \theta^{(h)}\right)}\right\}$.
    With probability α, set $\widehat{X}_i \leftarrow X_i'$; otherwise, retain the previous value.
  end for
  Step 2: Update parameters given $\widehat{X}$ (Gibbs step).
  Draw $\beta^{(h)}$ using (13).
  Draw $\gamma^{(h)}$ using (14).
  Increment h and return to Step 1.
end for

6.1. Numerical Implementation

In this section, we implement the methodologies outlined here by applying classical and MCMC methods to the **SIR** epidemic model, where the values of the model parameters γ and β are predefined. Using these predefined

values, we generate a sequence of N observations of the bivariate process using Gillespie's exact algorithm, as established in [14] (see Appendix A.3).

This implementation sets the parameters β and γ at values 0.2 and 0.5. The total population size is fixed at $n + a = 10^5$, with $a = 10$. The time increment is $\Delta t = \frac{T}{N}$, where the number of observations (including latent data) is given by $N = T \times M$, with $T = 60$ representing the simulation horizon and M representing the number of augmenting data points; $M \in \{2, 5, 10\}$. The initial values of the hyperparameters for the priors of β and γ are chosen following the guidelines in [12, Sect.5]. In order to address computational resource restrictions, we perform 25,000 iterations of the MCMC sampler for each dataset, under three alternative settings of the latent data points: $M = 2, 5$, and 10.

From a computational perspective, the cost of the MCMC algorithm increases with the number of latent points M, since a finer temporal discretization requires updating a larger number of augmented states at each iteration. In our implementation, the runtime grows approximately linearly with M. However, the numerical results show that the posterior summaries of (β, γ) stabilize for moderate values of M, indicating that increasing M beyond this range yields only marginal gains in estimation accuracy at a substantially higher computational cost. Consequently, moderate values of M offer a practical trade-off between statistical efficiency and computational feasibility.

The subfigures in Figure 1 and Figure 2 (first row) present the histograms of the ML estimates for β and γ in cases where $R_0 > 1$ and $R_0 \leq 1$, along with their corresponding normal probability density. The fitted normal distributions, shown, align closely with the empirical histograms of the estimates. This result suggests that the MLEs for both β and γ are approximately normally distributed, centered on their true values. This fact is provided by $Q - Q$ plots suggesting that the normality assumption is a real approximation for the distribution of the estimates, which is crucial for deriving valid confidence intervals.

Table 1 display the estimations of the parameters depending on M for three cases $R_0 > 1$, $R_0 = 1$ and $R_0 < 1$. As illustrated, the MLE's estimators exhibit less sensitivity to the number M, and the estimations remain close to the true values with slight deviations. However, the MCMC estimation method has an average sensitivity to the number M and provides more flexibility and robustness in handling missing data, which is crucial in practical applications where data may not be fully observed. However, this task requires more execution time than the MLE method. For comparison, commonly used approximate approaches such as deterministic ODE-based fitting or Gaussian increment methods were not considered here, as our focus is on exact discrete-time likelihood inference.

A summary representation of the posterior distribution is reported in Table 1 together with Figure 3. In particular, they report the posterior mean and the standard deviation. An increase in the number of latent data points M leads the estimates to converge towards the true values. However, the improvement is substantial only up to a moderate value of M, after which the incremental benefit becomes negligible.

As shown in Figure 3, the histograms, along with the MCMC trace plots displayed in Figure 4 and Figure 5, offer graphical evidence of the algorithm's convergence to the limiting distribution. In practice, when the true parameter values are unknown, increasing the number of latent points M may improve estimation precision, although this must be balanced against the associated computational cost [13].

6.2. Real Application

In this section, we apply both proposed estimation methods for the recovery and infection rates, γ and β, using COVID-19 data from Morocco for the period from 15$^{\text{th}}$ May 2021 to 30$^{\text{th}}$ November 2021, a total of 200 daily observations. This database is sourced from the World Health Organization [34], the United Nations Office for the Coordination of Humanitarian Affairs [36], and the official Moroccan coronavirus portal [35], which contains data on new and cumulative deaths as well as new and cumulative cases. The number M of latent points simulated between two successive periods, as explained in Section 4, is $M = 2, 5, 10$. The results are shown in Table 2.

By analyzing the period considered in this study, as shown in Table 2, the estimation of the parameters β and γ for both the classical and Bayesian approaches yields similar values. However, the credible interval obtained using the proposed MCMC algorithm is more accurate than the confidence interval obtained using the maximum likelihood method. This consistency across methodologies indicates robustness in the estimation process. Notably, the ratio $\frac{\beta}{\gamma}$, representing the effective reproduction number R_0, is very close to the values reported in [33] or [12].

Table 1. Means, standard deviations (SD), confidence intervals (CI) for $\hat{\beta}_{ML}$ and $\hat{\gamma}_{ML}$, and credible intervals for $\hat{\beta}_{MCMC}$ and $\hat{\gamma}_{MCMC}$, at different M values.

	β	$\hat{\beta}_{MCMC}$	$\hat{\beta}_{ML}$	γ	$\hat{\gamma}_{MCMC}$	$\hat{\gamma}_{ML}$
$M=2$						
mean	0.5	0.50416	0.50679	0.2	0.19751	0.19853
SD		0.00621	0.00321		0.00244	0.00126
CI(95%)		[0.4938, 0.5147]	[0.50049, 0.51309]		[0.1934, 0.2016]	[0.19607, 0.20100]
$M=5$						
mean	0.5	0.50324	0.50671	0.2	0.19778	0.19855
SD		0.00580	0.00335		0.00254	0.00192
CI(95%)		[0.4941, 0.5151]	[0.50049, 0.51309]		[0.1937, 0.2035]	[0.19663, 0.20047]
$M=10$						
mean	0.5	0.50245	0.50540	0.2	0.19801	0.19930
SD		0.00573	0.00218		0.00250	0.00141
CI(95%)		[0.4949, 0.5139]	[0.50321, 0.50763]		[0.1934, 0.2016]	[0.19791, 0.20072]
$M=2$						
mean	0.5	0.50716	0.50654	0.5	0.0.495421	0.50154
SD		0.00629	0.00331		0.00244	0.00126
CI(95%)		[0.4928, 0.5152]	[0.50056, 0.51311]		[0.48298, 0.51231]	[0.48674, 0.51204]
$M=5$						
mean	0.5	0.51357	0.50971	0.5	0.50916	0.50836
SD		0.00911	0.00428		0.00241	0.00182
CI(95%)		[0.4873, 0.5406]	[0.49412, 0.52088]		[0.4686, 0.5156]	[0.50746, 0.50927]
$M=10$						
mean	0.5	0.50449	0.50564	0.5	0.508601	0.50894
SD		0.00582	0.00284		0.00351	0.00312
CI(95%)		[0.4941, 0.5141]	[0.50340, 0.50771]		[0.4929, 0.52140]	[0.49801, 0.52040]
$M=2$						
mean	0.2	0.21051	0.20941	0.5	0.49510	0.0.49581
SD		0.00131	0.00129		0.00244	0.00126
CI(95%)		[0.18519, 0.21307]	[0.50049, 0.51309]		[0.48984, 0.51842]	[0.48059, 0.51005]
$M=5$						
mean	0.2	0.20152	0.20480	0.5	0.50746	0.50646
SD		0.00111	0.0012		0.00212	0.00183
CI(95%)		[0.4873, 0.5406]	[0.49444, 0.52198]		[0.4696, 0.5142]	[0.50784, 0.51027]
$M=10$						
mean	0.2	0.20112	0.20245	0.5	0.50881	0.50901
SD		0.00181	0.00169		0.00359	0.00362
CI(95%)		[0.19912, 0.21041]	[0.19941, 0.21101]		[0.4989, 0.52151]	[0.49824, 0.52068]

Table 2. Estimates of the parameters β and γ for data from Morocco between 15^{th} May 2021 and 30^{th} November 2021, with 25,000 iterations and $M \in \{2, 5, 10\}$, including their credible and confidence intervals.

	$\hat{\beta}_{MCMC}$	$\hat{\gamma}_{MCMC}$	$\hat{R}_{0,MCMC}$	$\hat{\beta}_{ML}$	$\hat{\gamma}_{ML}$	$\hat{R}_{0,ML}$
$M=2$						
mean	0.50506	0.12802	3.94516	0.49991	0.12731	3.92671
sd	0.00631	0.00251		0.00301	0.00125	
CI(95%)	[0.4907, 0.5188]	[0.1078, 0.1511]		[0.4812, 0.5130]	[0.1089, 0.1421]	
$M=5$						
mean	0.50373	0.12741	3.95361	0.50221	0.12699	3.95472
sd	0.00564	0.00241		0.00312	0.00182	
CI(95%)	[0.4917, 0.5188]	[0.1092, 0.1424]		[0.5041, 0.5109]	[0.1014, 0.1408]	
$M=10$						
mean	0.50112	0.12698	3.94644	0.50138	0.12694	3.94974
sd	0.00554	0.00250		0.00259	0.00146	
CI(95%)	[0.4948, 0.5210]	[0.1124, 0.1329]		[0.50211, 0.5125]	[0.1104, 0.1382]	

Figure 6 shows that the posterior predictive median trajectories for $I(t)$ and $R(t)$ are in good agreement with the observed Moroccan COVID-19 data. The associated 95% credible intervals capture the main epidemic phases, supporting the adequacy of the proposed SIR–Skellam model for real data analysis.

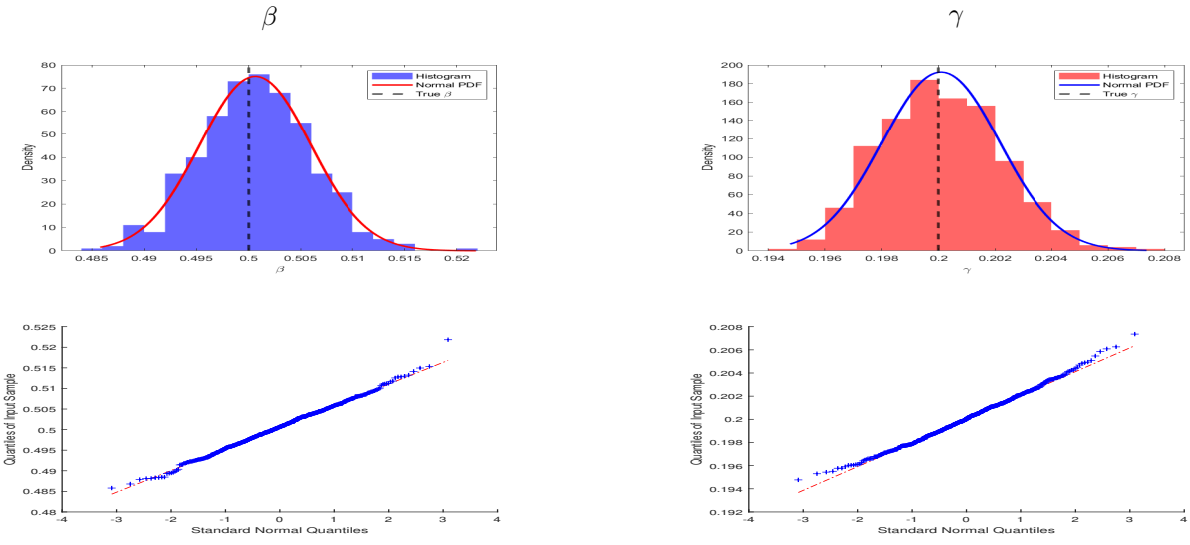

Figure 1. Histograms and the Q-Q plots for ML estimates of β (left column) and γ (right column) in the case where $R_0 > 1$. (For clarification on color references in the figure legend, consult the electronic version of this article.)

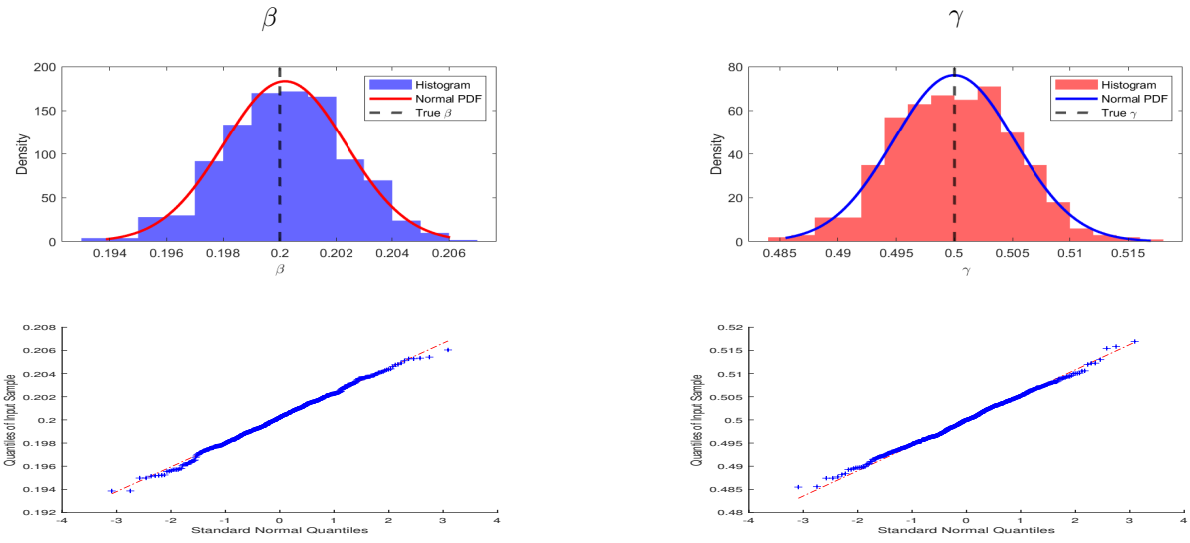

Figure 2. Histograms and the Q-Q plots for ML estimates for β (left column) and γ (right column) in the case where $R_0 \leq 1$. (For clarification on color references in the figure legend, consult the digital version of this article.)

7. Conclusion

In this study, we explored both classical and Bayesian methods for estimating the infection and recovery parameters, β and γ, in a stochastic SIR epidemic model, with particular attention to the implementation of the Skellam distribution [30]. The Skellam distribution proved to be a robust and flexible tool for modeling the dynamics of epidemic processes, capturing the interaction between infection and recovery events in a closed and homogeneous population.

The present work shows the strengths and limitations of both the maximum likelihood and Bayesian estimation approaches. While the classical maximum likelihood method provided consistent and computationally efficient estimates, the Bayesian framework offered greater flexibility, especially in scenarios with sparse or incomplete

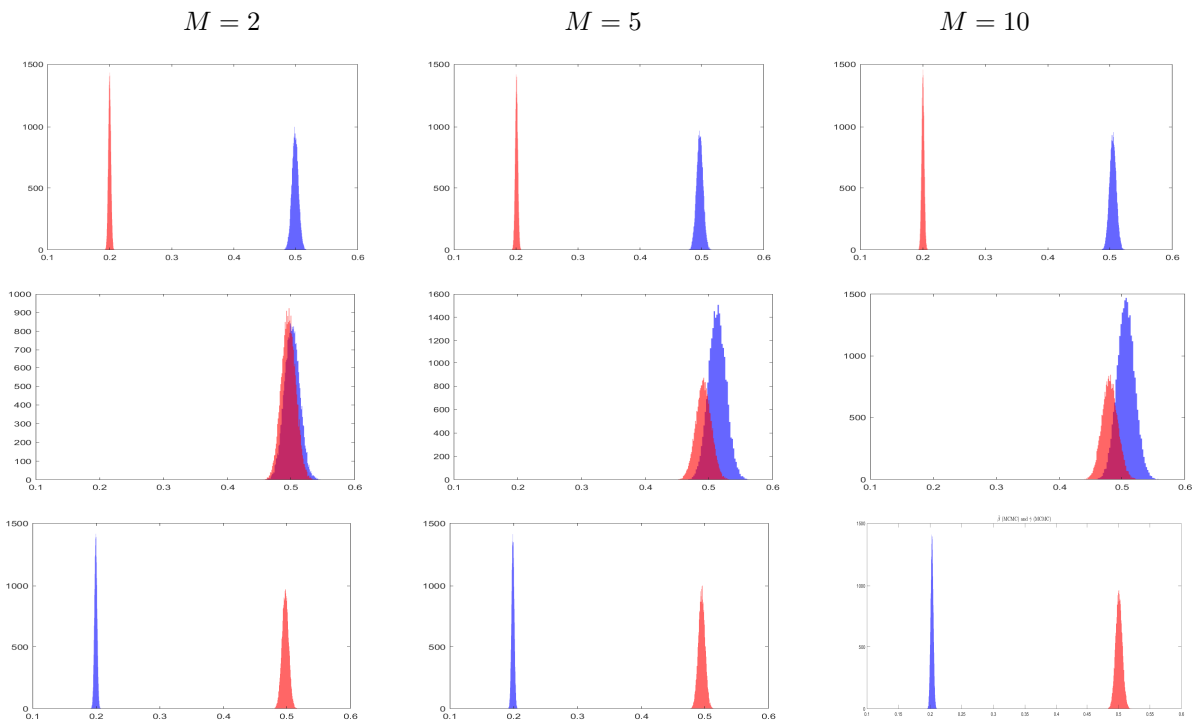

Figure 3. Frequency histograms of the MCMC chains with 25,000 iterations for the posterior density estimates of $\hat{\beta}_{MCMC}$ (blue) and $\hat{\gamma}_{MCMC}$ (red) for $M = 2, 5, 10$ across the three cases: $\hat{R}_0 > 1$ (first row), $\hat{R}_0 = 1$ (second row), and $\hat{R}_0 < 1$ (third row). (For clarification on color references in the figure legend, consult the electronic version of this article.)

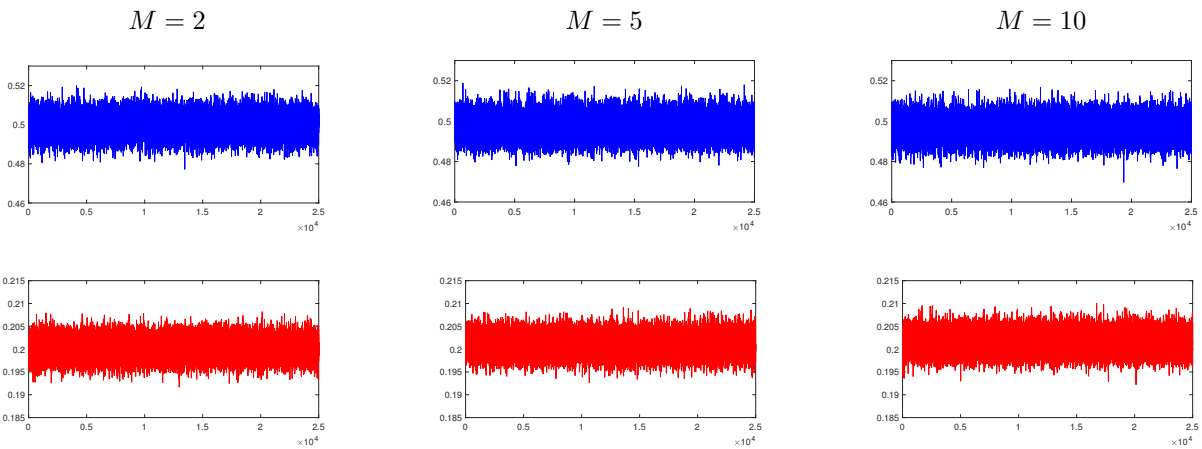

Figure 4. Trace plots of the MCMC chains based on 25,000 iterations for the posterior density estimates of $\hat{\beta}_{MCMC}$ (first row) and $\hat{\gamma}_{MCMC}$ (second row) for the case where $\hat{R}_0 > 1$.

data. By incorporating prior knowledge through independent Gamma distributions, which serve as an appropriate selection for positive parameters. The Bayesian approach has significantly improved estimation accuracy, as demonstrated by numerical simulation.

The application of the proposed methods to real COVID-19 data from Morocco highlighted the practical relevance of our approach. The results obtained are consistent when compared with existing research and

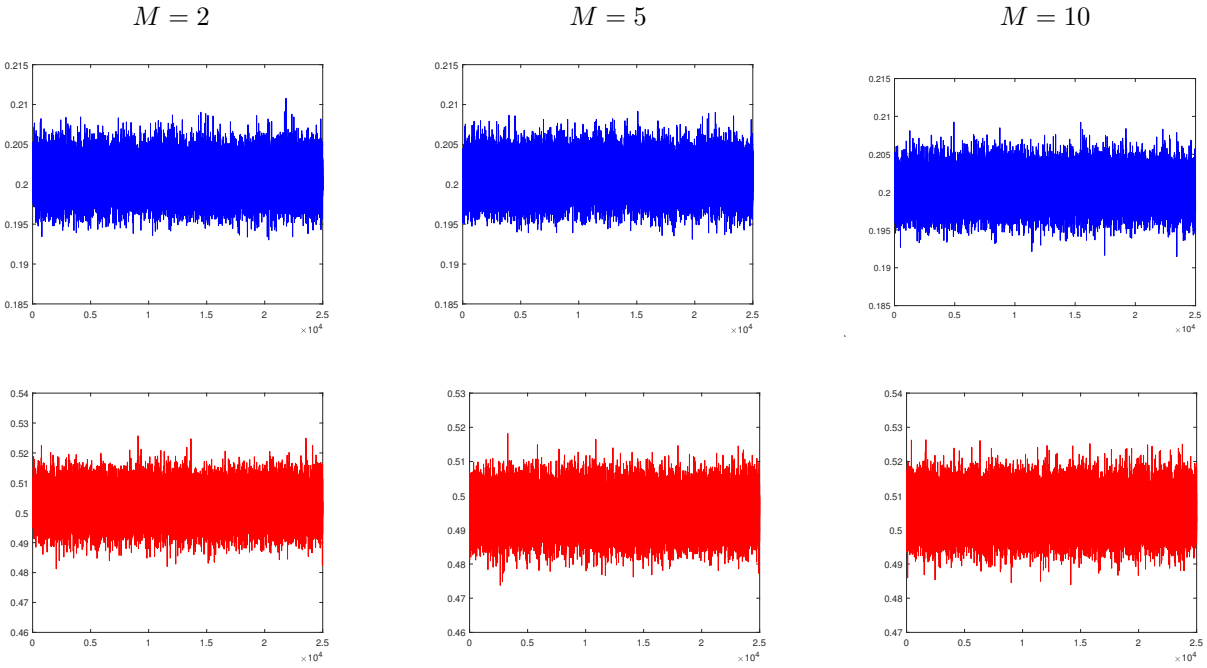

Figure 5. Trace plots of the MCMC chains based on 25,000 iterations for the posterior density estimates of $\hat{\beta}_{MCMC}$ (first row) and $\hat{\gamma}_{MCMC}$ (second row) for the case where $\hat{R}_0 \leq 1$.

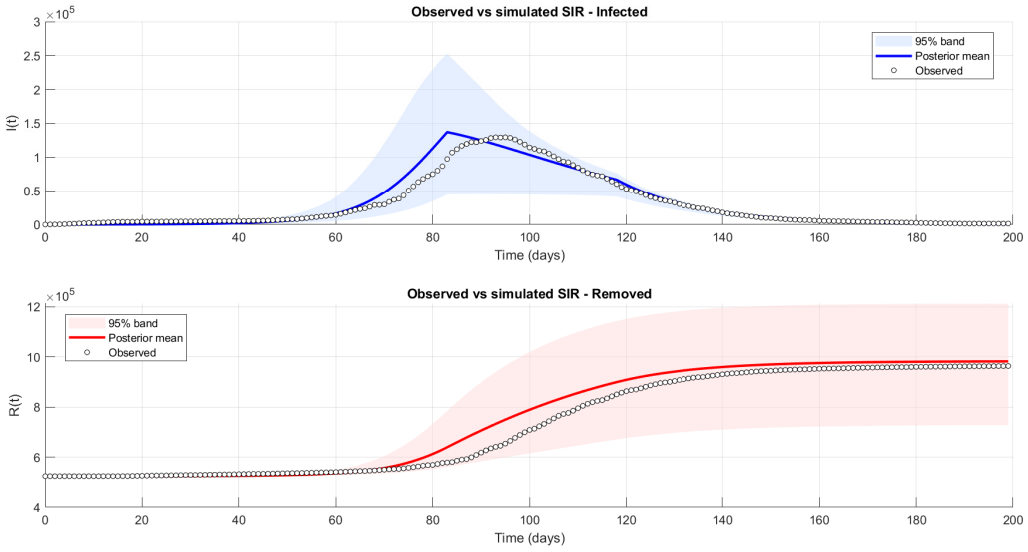

Figure 6. Observed Moroccan COVID-19 data (black circles) with posterior median trajectories and 95% credible intervals for $I(t)$ (top) and $R(t)$ (bottom) simulated from the posterior distribution.

demonstrate the robustness of the proposed methodologies, e.g., [12]. The Bayesian method showed its potential to address challenges inherent to real epidemic data, such as variability and missing observations.

Looking forward, the methodologies and findings of this study can be extended to more complex epidemic models, such as those incorporating heterogeneous populations or additional compartments, including SEIR-type

structures with time-varying transmission rates. In addition, future work will investigate systematic comparisons between the proposed Skellam-based MLE and Bayesian estimators and commonly used inference approaches, such as deterministic ODE-based estimators or Gaussian approximation methods, in order to further assess their relative performance under different data regimes. Moreover, the integration of additional sources of uncertainty, such as measurement errors or unobserved heterogeneity, could further enhance the applicability of these approaches in real-world epidemiological settings.

Acknowledgements

The authors would like to thank the editors and the anonymous referees for their helpful comments and suggestions, which contributed to improving the quality of this manuscript.

Declaration of conflicting interests

The authors declare that no potential conflicts of interest, financial or non-financial, exist in relation to the content of this article.

A. Appendix

A.1. The proof of theorem (1)

For ease of notation, we define $a_i = \int_{t_{i-1}}^{t_i} I(s)\big(a+n-I(s)-R(s)\big)\,ds$ and $b_i = \int_{t_{i-1}}^{t_i} I(s)\,ds$. By applying Bayes' rule, we obtain

$$\begin{aligned}
\pi\Big(X_i=(x,y)\mid X_{i+1},X_{i-1},\beta,\gamma\Big) &\propto \mathbb{P}\Big(X_i=(x,y),X_{i+1},X_{i-1}\mid\beta,\gamma\Big)\\
&\propto \mathbb{P}\left(X_i=(x,y)\mid X_{i-1},\beta,\gamma\right)\times\mathbb{P}\left(X_{i+1}\mid X_i,\beta,\gamma\right)\\
&\propto \exp\left\{-\beta\frac{a_i}{n}-\gamma b_i\right\}\frac{\left(\beta\frac{a_i}{n}\right)^{x+y-a}}{(x+y-a)!}\times\frac{(\gamma b_i)^y}{y!}\\
&\quad \exp\left\{-\beta\frac{a_{i+1}}{n}-\gamma b_{i+1}\right\}\frac{\left(\beta\frac{a_{i+1}}{n}\right)^{I_{t_{i+1}}+R_{t_{i+1}}-a}}{(I_{t_{i+1}}+R_{t_{i+1}}-a)!}\times\frac{(\gamma b_{i+1})^{R_{t_{i+1}}}}{R_{t_{i+1}}!}\\
&\propto \exp\left\{-\beta\frac{(a_i+a_{i+1})}{n}-\gamma(b_i+b_{i+1})\right\}\left(\frac{a_{i+1}}{n}\right)^{I_{t_{i+1}}+R_{t_{i+1}}-a}\\
&\quad (b_{i+1})^{R_{t_{i+1}}}\frac{\beta^{I_{t_{i+1}}+x+R_{t_{i+1}}+y-2a}.\gamma^{y+R_{t_{i+1}}}}{(x+y-a)!}\left(\frac{a_i}{n}\right)^{x+y-a}b_i^y\\
&\propto \exp\left\{-\beta\frac{(a_i+a_{i+1})}{n}-\gamma(b_i+b_{i+1})\right\}\left(\frac{a_{i+1}}{n}\right)^{I_{t_{i+1}}+R_{t_{i+1}}-a}\\
&\quad (b_{i+1})^{R_{t_{i+1}}}\frac{\beta^{x+y}.\gamma^y}{(x+y-a)!y!}\left(\frac{a_i}{n}\right)^{x+y-a}b_i^y. \qquad (24)
\end{aligned}$$

In order to obtain (12), we approximate the integrals a_i, a_{i+1}, b_i, and b_{i+1} by using the first-order trapezoidal method;

$$\begin{aligned}\frac{a_i}{n} &= \frac{1}{n}\int_{t_{i-1}}^{t_i} I(s)\,(n+a-I(s)-R(s))\,ds \\ &= \frac{\delta t}{2n}x(a+n-x-y)+\underbrace{\frac{\delta t}{2n}\left[I_{t_{i-1}}(a+n-I_{t_{i-1}}-R_{t_{i-1}})\right]}_{C_1^i},\end{aligned}$$

and

$$\begin{aligned}\frac{a_{i+1}}{n} &= \frac{1}{n}\int_{t_i}^{t_{i+1}} I(s)\,(n+a-I(s)-R(s))\,ds \\ &= \frac{\delta t}{2n}\big(x(a+n-x-y)\big)+\underbrace{\frac{\delta t}{2n}\left(I_{t_{i+1}}(a+n-I_{t_{i+1}}-R_{t_{i+1}})\right)}_{C_2^i}.\end{aligned}$$

Furthermore

$$\begin{aligned}\frac{a_i+a_{i+1}}{n} &= \frac{1}{n}\left(\int_{t_{i-1}}^{t_i} I(s)(a+n-I(s)-R(s))ds+\int_{t_i}^{t_{i+1}} I(s)(n+a-I(s)-R(s))ds\right) \\ &= \frac{\delta t}{n}\left(I_{t_{i-1}}(n+a-I_{t_{i-1}}-R_{t_{i-1}})+I_{t_{i+1}}(n+a-I_{t_{i+1}}-R_{t_{i+1}})\right) \\ &= 2(C_1^i+C_2^i),\end{aligned}$$

and

$$\begin{aligned}b_i &= \int_{t_{i-1}}^{t_i} I(s)ds = \frac{t_i-t_{i-}}{2}\left(I_{t_{i-1}}+I_{t_i}\right) \\ &\simeq \frac{\delta t}{2}\left(x+I_{t_{i-1}}\right) \\ &= \frac{\delta t}{2}x+\underbrace{\frac{\delta t}{2}I_{t_{i-1}}}_{C_3^i},\end{aligned}$$

and

$$\begin{aligned}b_{i+1} &= \int_{t_i}^{t_{i+1}} I(s)ds = \frac{t_{i+1}-t_i}{2}\left(I_{t_i}+I_{t_{i+1}}\right) \\ &\simeq \frac{\delta t}{2}\left(x+I_{t_{i+1}}\right) \\ &= \frac{\delta t}{2}x+\underbrace{\frac{\delta t}{2}I_{t_{i+1}}}_{C_4^i},\end{aligned}$$

and

$$\begin{aligned}b_i+b_{i+1} &= \int_{t_{i-1}}^{t_{i+1}} I(s)ds \simeq \frac{t_{i+1}-t_{i-1}}{2}\left[I_{t_{i+1}}+I_{t_{i-1}}\right] \\ &= \delta t\left(I_{t_{i-1}}+I_{t_{i+1}}\right) \\ &= 2(C_3^i+C_4^i),\end{aligned}$$

we set $C_5^i = I_{t_{i+1}} + R_{t_{i+1}} - a$.
Then by inserting those previously terms in (24), we found that

$$\begin{aligned}\pi\Big(X_i = (x,y) \mid X_{i-1}, X_{i+1}, \beta, \gamma\Big) \propto \exp\left\{-2\beta(C_1^i + C_2^i) - 2\gamma(C_3^i + C_4^i)\right\} \\ \left(\frac{\delta t}{2n} x(n+a-x-y) + C_1^i\right)^{x+y-a} \\ \left(\frac{\delta t}{2n} x(n+a-x-y) + C_2^i\right)^{C_5^i} \left(\frac{\delta t}{2} x + C_4^i\right)^{y} \\ \left(\frac{\delta t}{2} x + C_5^i\right)^{R_{t_{i+1}}} \frac{\beta^{x+y}\gamma^y}{(x+y-a)!y!}.\end{aligned}$$

Then

$$\begin{aligned}\pi\Big(X_i = (x,y) \mid X_{i-1}, X_{i+1}, \beta, \gamma\Big) \propto \left(\frac{\delta t}{2n} x(n-x-y) + C_1^i\right)^{x+y-a} \\ \left(\frac{\delta t}{2n} x(n-x-y) + C_2^i\right)^{C_5^i} \left(\frac{\delta t}{2} x + C_3^i\right)^{y} \\ \left(\frac{\delta t}{2} x + C_4^i\right)^{R_{t_{i+1}}} \frac{\beta^{x+y}\gamma^y}{(x+y-a)!y!},\end{aligned}$$

where

$$\begin{array}{lll} C_1^i = \frac{\delta t}{2n}\left[I_{t_{i-1}}(n+a-I_{t_{i-1}}-R_{t_{i-1}})\right] & \text{and} & C_2^i = \frac{\delta t}{2n}\left[I_{t_{i+1}}(n+a-I_{t_{i+1}}-R_{t_{i+1}})\right] \\ C_3^i = \frac{\delta t}{2} I_{t_{i-1}} & \text{and} & C_4^i = \frac{\delta t}{2} I_{t_{i+1}}. \end{array} \tag{25}$$

Hence, the proof of (12) in Theorem (1) is concluded.

A.2. The proof of theorem (2)

By applying the formula (9) and also by the same notations as used previously for a_i and b_i. The likelihood for $\theta = (\beta, \gamma)$ cab written as:

$$
\begin{aligned}
L(\hat{\mathbf{X}} \mid \theta) &= \prod_{i=1}^{N} \mathbb{P}\big(\hat{X}_i = (I_{t_i}, R_{t_i}) \mid \hat{X}_{i-1}, \theta\big) \\
&= \prod_{i=1}^{N} \exp\left(-\lambda_1^i - \lambda_2^i\right) \frac{(\lambda_1^i)^{I_{t_i}+R_{t_i}-a}}{(I_{t_i}+R_{t_i}-a)!} \frac{(\lambda_2^i)^{R_{t_i}}}{R_{t_i}!} \\
&= \prod_{i=1}^{N} \exp\left(-\frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)(a+n-I(s)-R(s))ds - \gamma\int_{t_{i-1}}^{t_i} I(s)ds\right) \\
&\quad \times \left[\frac{\beta}{n}\int_{t_{i-1}}^{t_i} I(s)(a+n-I(s)-R(s))ds\right]^{I_{t_i}+R_{t_i}-a} \times \left[\gamma\int_{t_{i-1}}^{t_i} I(s)ds\right]^{R_{t_i}} \\
&= \exp\left(-\frac{\beta}{n}\left[\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i} I(s)(a+n-I(s)-R(s))ds\right] - \gamma\left[\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i} I(s)ds\right]\right) \\
&\quad \beta^{\left(\sum_{i=1}^{N}(I_{t_i}+R_{t_i}-a)\right)} \gamma^{\left(\sum_{i=1}^{N} R_{t_i}\right)} \\
&= \exp\left(-\beta \underbrace{\frac{1}{n}\left[\sum_{i=1}^{N} a_i\right]}_{A_{1,N}} - \gamma \underbrace{\left[\sum_{i=1}^{N} b_i\right]}_{A_{2,N}}\right) \times \beta^{\overbrace{\left(\sum_{i=1}^{N}(I_{t_i}+R_{t_i}-a)\right)}^{A_{3,N}}} \times \gamma^{\overbrace{\left(\sum_{i=1}^{N} R_{t_i}\right)}^{A_{4,N}}},
\end{aligned}
$$

where the constants $A_{1,N}, \cdots, A_{4,N}$ are previously defined in (15).

Then

$$
L\left(\hat{\mathbf{X}} \mid \theta\right) \propto \exp\left(-\beta A_{1,N} - \gamma A_{2,N}\right) \times \beta^{A_{3,N}} \gamma^{A_{4,N}}. \tag{26}
$$

According to Bayes theorem, we have $\pi(\beta \mid \hat{\mathbf{X}}, \gamma) \propto L\left(\hat{\mathbf{X}} \mid \theta\right) \times \pi(\theta)$, and based on the fact that β and γ are two parameters priorly independents, and by exploring (26) we get

$$
\begin{aligned}
\pi\left(\beta \mid \hat{\mathbf{X}}, \gamma\right) &\propto L\left(\hat{\mathbf{X}} \mid \theta\right) \times \pi(\theta) \\
&\propto L\left(\hat{\mathbf{X}} \mid \beta, \gamma\right) \times \pi(\beta)\pi(\gamma) \\
&\propto \exp(-\beta A_{1,N} - \gamma A_{2,N}) \times \beta^{A_{3,N}} \pi(\beta)\pi(\gamma) \\
&\propto \exp(-\beta A_{1,N}) \beta^{A_{3,N}} \pi(\beta).
\end{aligned}
$$

Furthermore, since β follows a Gamma prior with parameters (m, λ), it follows that

$$
\begin{aligned}
\pi\left(\beta \mid \hat{\mathbf{X}}, \gamma\right) &\propto \exp(-\beta A_{1,N}) \beta^{A_{3,N}} \times \beta^{m-1} \times \exp\left(-\lambda\beta\right) \\
&\propto \exp\left(-\beta\left(A_{1,N} + \lambda\right)\right) \beta^{A_{3,N}+m-1} \\
&\propto \Gamma\left(A_{3,N} + m - 1, A_{1,N} + \lambda\right).
\end{aligned}
$$

Finally, we obtain the proof of formula (13).
We proceed by the same demonstration as above, we have

$$\begin{aligned}\pi\left(\gamma \mid \hat{\mathbf{X}}, \beta\right) &\propto L\left(\hat{\mathbf{X}} \mid \theta\right) \times \pi(\theta) \\ &\propto \exp(-\beta A_{1,N} - \gamma A_{2,N}) \beta^{A_{3,N}} \gamma^{A_{4,N}} \pi(\beta)\pi(\gamma) \\ &\propto \exp(-\gamma A_{2,N}) \gamma^{A_{4,N}} \pi(\gamma).\end{aligned}$$

Given that γ is assumed to follow a Gamma prior with parameters (m', λ'), we obtain

$$\begin{aligned}\pi\left(\gamma \mid \hat{\mathbf{X}}, \beta\right) &\propto \exp(-\gamma A_{2,N}) \times \gamma^{A_{4,N}} \times \pi(\gamma) \\ &\propto \exp\left(-\gamma(A_{2,N} + \lambda')\right) \gamma^{A_{4,N}+m'-1} \\ &\propto \Gamma\left(A_{4,N} + m' - 1, A_{2,N} + \lambda'\right).\end{aligned}$$

We obtain the proof of the formula (14).

A.3. The Gillespie algorithm

As described in [14], Gillespie's algorithm is an exact procedure to simulate continuous-time pure Markov jump processes. For the SIR framework, the state of the population at time t can be written as $(S(t), I(t), R(t)) = (s, i, a + n - s - i)$, where s is the number of susceptibles, i is the number of infected, and n is the total population size.
The two events considered in the SIR model are infection with rate $\alpha_1 = \beta \frac{si}{n}$ and recovery with rate $\alpha_2 = \gamma i$, where $\mathcal{E}(.)$ will denote the exponential distribution. The algorithm is implemented in two main steps: This algorithm

Algorithm 2 Gillespie Algorithm for SIR Model

Initialize: $t \leftarrow 0$, $S(0) \leftarrow s_0$, $I(0) \leftarrow i_0$, $R(0) \leftarrow n - s_0 - i_0$
while $t < t_{\max}$ **do**
  Calculate rates: $\alpha_1 \leftarrow \beta \frac{I(t)(a+n-I(t)-R(t))}{n}$, $\alpha_2 \leftarrow \gamma I(t)$, $\alpha \leftarrow \alpha_1 + \alpha_2$
  Draw τ from $\mathcal{E}(\alpha)$
  Choose k according to probabilities $p_k = \frac{\alpha_k}{\alpha}$
  if $k = 1$ **then**
    Infection event: $S \leftarrow S - 1$, $I \leftarrow I + 1$, $R \leftarrow R$
  else
    Recovery event: $S \leftarrow S$, $I \leftarrow I - 1$, $R \leftarrow R + 1$
  end if
  Update time: $t \leftarrow t + \tau$
end while

captures the stochastic transitions between the compartments of the SIR model. The exponential waiting times and event probabilities ensure that the simulation accurately represents the continuous-time Markov jump process associated with the SIR epidemic dynamics.

REFERENCES

1. A. Abou-Bakre, Z. Boufis, H. El Maroufy, and T. Kernane, *Bayesian inference of a stochastic diffusion process for the dynamic of HIV in closed heterosexual population with simulations and application to Morocco case*, Appl. Math. Model., vol. 106, pp. 448–462, 2022.
2. M. Abramowitz, I. A. Stegun, and R. H. Romer, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Am. J. Phys., vol. 56, no. 10, p. 958, 1988.
3. M. F. Alharthi, *The Basic Reproduction Number for the Markovian SIR-Type Epidemic Models: Comparison and Consistency*, J. Math., vol. 2022, 2022.

4. A. A. Alzaid and M. A. Omair, *On the Poisson difference distribution inference and applications*, Bull. Malays. Math. Sci. Soc. (2), 33(1):17–45, 2010.
5. H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, Springer Science and Business Media, New York, 2012.
6. F. Ball and P. O'Neill, *A modification of the general stochastic epidemic motivated by AIDS modelling*, Adv. Appl. Probab., vol. 25, no. 1, pp. 39–62, 1993.
7. N. T. J. Bailey, *The mathematical theory of infectious diseases and its applications. No. 2nd edition*, 1975.
8. M. J. Crowder, *Maximum likelihood estimation for dependent observations*, J. R. Stat. Soc. Ser. B, vol. 38, no. 1, pp. 45–53, 1976.
9. H. El Maroufy, T. Kernane, S. Becheket, and A. Ouddadj, *Bayesian inference for nonlinear stochastic SIR epidemic model*, J. Stat. Comput. Simul., vol. 86, no. 11, pp. 2229–2240, 2016.
10. H. El Maroufy, L. Omari, and Z. Taib, *Transition probabilities for generalized SIR epidemic model*, Stoch. Models, vol. 28, no. 1, pp. 15–28, 2012.
11. S. R. Eliason, *Maximum likelihood estimation: logic and practice*, Sage, Thousand Oaks, CA, vol. 96, 1993.
12. H. El Maroufy and A. Lagzini, *Bayesian inference for SIS type epidemic model by Skellam's distribution with real application to COVID-19*, Stat. Biosci., pp. 1–26, 2024.
13. B. Eraker, *MCMC Analysis of Diffusion Models With Application to Finance*, J. Bus. Econ. Stat., vol. 19, no. 2, pp. 177–191, 2001.
14. M. Gibson and J. Bruck, *An efficient algorithm for generating trajectories of stochastic gene regulation reactions*, Caltech Parallel Distrib. Syst. Group, Tech. Rep., 1998.
15. J. O. Irwin, *The frequency distribution of the difference between two independent variates following the same Poisson distribution*, J. R. Stat. Soc. Ser. A Stat. Soc., vol. 100, no. 3, pp. 415–416, 1937.
16. D. Karlis and I. Ntzoufras, *Analysis of sports data by using bivariate Poisson models*, J. R. Stat. Soc. Ser. D, vol. 52, no. 3, pp. 381–393, 2003.
17. S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, 2009.
18. D. F. Anderson, A. Ganguly, and T. G. Kurtz, *Error analysis of tau-leap simulation methods*, Ann. Appl. Probab., vol. 21, no. 6, pp. 2226–2262, 2011.
19. M. Martcheva, *An introduction to mathematical epidemiology*, Springer, vol. 61, 2015.
20. J. R. Norris, *Markov chains*, Cambridge University Press, 2nd edition, 1998.
21. P. D. O'Neill and G. O. Roberts, *Bayesian inference for partially observed stochastic epidemics*, J. R. Stat. Soc. Ser. A, vol. 162, pp. 121–129, 1999.
22. A. Prékopa, *On composed Poisson distributions, IV*, Acta Math. Acad. Sci. Hungar., vol. 3, pp. 317–325, 1952.
23. J.-X. Pan and K.-T. Fang, *Maximum likelihood estimation*, Growth Curve Mod. Stat. Diagn., pp. 77–158, 2002.
24. A. Qaffou, H. El Maroufy, and T. Kernane, *Bayesian parameter inference for stochastic SIR epidemic model with hyperbolic diffusion*, Commun. Stat. Simul. Comput., vol. 46, no. 9, pp. 6907–6922, 2017.
25. C. P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, 2nd edition, New York, 1999.
26. C. Robert, *Le choix bayésien: principes et pratique*, Springer Science and Business Media, New York, 2005.
27. S. M. A. Aljeddani and M. A. Mohammed, *Parameter estimation of a model using maximum likelihood function and Bayesian analysis through moment of order statistics*, Alex. Eng. J., vol. 75, pp. 221–232, 2023.
28. A. N. Shiryaev, *Probability-1*, Springer, New York, 2016.
29. M. J. Schervish, *Theory of statistics*, Springer Science and Business Media, 2012.
30. J. G. Skellam, *The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations*, J. R. Stat. Soc. Ser. A, vol. 109, no. 3, pp. 109–296, 1946.
31. T. J. Sweeting, *Uniform asymptotic normality of the maximum likelihood estimator*, Ann. Stat., pp. 1375–1381, 1980.
32. L. Tierney, *Markov chains for exploring posterior distributions*, Ann. Stat., vol. 22, no. 4, pp. 1701–1728, 1994.
33. Tableau Public, *Global COVID-19 Reproduction Number and Doubling Time Tracker*, Retrieved from https://tinyurl.com/5n7b4kp2, 2023.
34. WHO, *WHO Coronavirus (COVID-19) Dashboard*, Retrieved from https://covid19.who.int/region/emro/country/ma, 2023.
35. POCM, *Le Portail Officiel du Coronavirus au Maroc*, Retrieved from http://www.covidmaroc.ma/Pages/Accueilfr.aspx, 2023.
36. OCHA, *United Nations Office for the Coordination of Humanitarian Affairs*, Retrieved from https://tinyurl.com/2jkryolu, 2023.