Approximation of epidemic models by diffusion processes and their statistical inference

Romain Guy · Catherine Larédo · Elisabeta Vergu

Received: 15 May 2013 / Revised: 6 March 2014 / Published online: 27 March 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Multidimensional continuous-time Markov jump processes \((Z(t))\) on \(\mathbb{Z}^p\) form a usual set-up for modeling SIR-like epidemics. However, when facing incomplete epidemic data, inference based on \((Z(t))\) is not easy to be achieved. Here, we start building a new framework for the estimation of key parameters of epidemic models based on statistics of diffusion processes approximating \((Z(t))\). First, previous results on the approximation of density-dependent SIR-like models by diffusion processes with small diffusion coefficient \(1/\sqrt{N}\), where \(N\) is the population size, are generalized to non-autonomous systems. Second, our previous inference results on discretely observed diffusion processes with small diffusion coefficient are extended to time-dependent diffusions. Consistent and asymptotically Gaussian estimates are obtained for a fixed number \(n\) of observations, which corresponds to the epidemic context, and for \(N \to \infty\). A correction term, which yields better estimates non asymptotically, is also included. Finally, performances and robustness of our estimators with respect to various parameters such as \(R_0\) (the basic reproduction number), \(N\), \(n\) are investigated on simulations. Two models, SIR and SIRS, corresponding to single and recurrent outbreaks, respectively, are used to simulate data. The findings indicate that our estimators have good asymptotic properties and behave noticeably well for realistic numbers of observations and population sizes. This study lays the foundations of a generic inference method currently under extension to incompletely observed epidemic data.

Electronic supplementary material The online version of this article (doi:10.1007/s00285-014-0777-8) contains supplementary material, which is available to authorized users.

R. Guy (ES) · C. Larédo · E. Vergu
UR 341 Mathématiques et Informatique Appliquées, INRA, Jouy-en-Josas, France
e-mail: romain.guy@jouy.inra.fr

R. Guy · C. Larédo
UMR 7599 Laboratoire de Probabilités et Modèles Aléatoires, Université Denis Diderot Paris 7 and CNRS, Paris, France
epidemic data. Indeed, contrary to the majority of current inference techniques for partially observed processes, which necessitates computer intensive simulations, our method being mostly an analytical approach requires only the classical optimization steps.

**Keywords** Diffusion approximation · Minimum contrast estimators · Time-dependent processes · Epidemics

**Mathematics Subject Classification (2000)** 62F12 · 60J27 · 60J60

1 Introduction

Mathematical modeling of epidemic spread and estimation of key parameters from data provides much insight in the understanding of public health problems related to infectious diseases. Classically, an epidemic dynamics in closed population of size \( N \) is described by the \( SIR \) model (Susceptible-Infectious-Removed from the infectious chain), where each individual can find himself at a given time in one of these three mutually exclusive health states. Systems with larger dimensionality can be obtained if the description of the epidemic dynamics is refined (Keeling and Rohani 2011). One of the most natural representations of the \( SIR \) model is a bidimensional continuous-time Markov jump process. Given that \( R(t) = N - S(t) - I(t) \) \( \forall t \), we can define \( Z(t) = (S(t), I(t)) \) with initial state \( Z(0) = (S_0, I_0) \) and transitions \( (S, I) \xrightarrow{\lambda} (S-1, I+1) \) and \( (S, I) \xrightarrow{\gamma} (S, I-1) \), where \( \lambda \) is the transmission rate and \( \gamma = 1/d \) is the recovery rate or the inverse of the mean infection duration \( d \). Beyond this process, the successive transitions of individuals between states were described in various mathematical frameworks: ODE/PDE (Diekmann and Heesterbeek 2000); difference equations and continuous or discrete-time stochastic processes (Daley and Gani 2001; Diekmann et al 2012), such as point processes, renewal processes, branching processes, diffusion processes.

These models are naturally parametric and allow estimating key parameters (such as transmission rate, mean sojourn time in the infectious state, extinction probability) through likelihood-based or M-estimation methods sometimes coupled to Bayesian methods (Andersson and Britton 2000). Due to the fact that epidemic data are most often partially observed (e.g. infectious and recovery dates are not observed for all individuals recorded in the surveillance system, not all the infectious individuals are reported) and also temporally and/or spatially aggregated, estimation through likelihood-based approaches is rarely straightforward, whatever the mathematical representation used. Various methods were developed during the last years to overcome this problem, data augmentation methods and likelihood-free methods being those which generated the keenest interest (Breto et al 2009; McKinley et al 2009). Although these methods allow considering different patterns of missingness, they do not provide a definitive solution to the statistical inference from epidemic data. Indeed, in practice there are some limitations due to the amount of augmented data and to the adjustment of the numerous tuning parameters (Andersson and Britton 2000; see also...
O’Neill (2010) for a short review of available statistical methods for relating models to data and for future challenges). Moreover, identifiability related issues are rarely addressed.

In this context, diffusion processes, which provide good approximations of epidemic dynamics (see e.g. Fuchs 2013; Ross et al 2009), allow shedding new light on inference related problems of epidemic data due to their analytical power. The normalization of the SIR Markov jump process \( (Z(t)) \) by \( N \) asymptotically leads to an ODE system \( x(t) = (s(t), i(t)) \) with \( s(0) = 1 - I(0)/N, i(0) = I(0)/N \) and 
\[
\frac{ds(t)}{dt} = -\lambda s(t)i(t) \quad \text{and} \quad \frac{di(t)}{dt} = \lambda s(t)i(t) - \gamma i(t).
\]
Before passing to the limit, one can describe the epidemic dynamics through a bidimensional diffusion with a small diffusion coefficient proportional to \( 1/\sqrt{N} \). On the statistical side, we can consider as a first approximation that epidemic data correspond to low frequency data (i.e. a fixed number of observations \( n \)) observed on a fixed interval \([0, T]\). Although this is an optimistic view of field data which are most often incomplete and incorporate observational errors, it lays the foundations of further investigations in more realistic contexts.

Historically, statistics for diffusions were developed for continuously observed processes which renders possible getting an explicit formulation of the likelihood (Kutoyants 1984; Lipster and Shiryaev 2001). As mentioned above, in practice, epidemic data are not continuous, but partial, with various mechanisms underlying the missingness and leading to intractable likelihoods: trajectories can be discretely observed with a sampling interval (low frequency or high frequency observations, i.e. \( n \to \infty \)); discrete observations can correspond to integrated processes; some coordinates can be unobserved. Since the nineties, statistical methods associated to the first two types of data have been developed (e.g. Genon-Catalot and Jacod 1993; Kessler 2000; Genon-Catalot et al 2000; Gloter 2001; Elerian et al 2001). Recently proposed approaches for multidimensional diffusions are based on the filtering theory (Genon-Catalot and Larédo 2006; Fearnhead et al 2008). Concerning diffusions with small diffusion coefficient from discrete observations, the asymptotic properties of estimators were largely studied over the two last decades (e.g. Larédo 1990; Genon-Catalot 1990; Genon-Catalot et al 2002; Sørensen and Uchida 2003; Gloter and Sørensen 2009) in various contexts (uni- and multidimensional cases, observations sampled at low and high frequency, discrete sampling of the state space). In this framework, it is important to distinguish drift parameters and diffusion parameters, because they are not estimated at the same rate. In a recent work (Guy et al 2014), we studied multidimensional diffusion with small variance, for discrete observations on both components of the system approximating \( (Z(t)) \). We provided minimum contrast estimators with good properties for both high and low-frequency observations on a fixed time interval \([0, T]\).

In this paper, we propose first, based on the results of Ethier and Kurtz (2005), a generic and rigorous method to construct multidimensional diffusion processes with small variance as mathematical representations of epidemic dynamics, by approximating a Markov jump process (Sect. 2). The approach of Ethier and Kurtz (2005) is extended to general density time-dependent Markov processes (Sect. 2.4). The second and main result is a new inference method for the parameters of the diffusion process obtained in Sect. 2, discretely observed (fixed \( n \)) on a fixed time interval and
for the special case where the same parameters are in the diffusion and drift terms. Building on the results of Guy et al (2014), we elaborate a new contrast based on the Gaussian approximation of the diffusion process (Sect. 3). In addition to consistent and asymptotically Gaussian minimum contrast estimators obtained for fixed $n$ (which corresponds to the epidemic context) and for $N \to \infty$, the correction term we introduce in the new contrast allows yielding better estimates non asymptotically. We also extend the results of Guy et al (2014) to time-dependent diffusions. Finally, the accuracy of these estimators is explored on simulated epidemic data (Sect. 4) for single outbreaks ($SIR$) and for recurrent epidemics (non autonomous $SIRS$, i.e. with seasonal forcing in transmission). The case of discrete observations of all the coordinates that we investigate in this study also provides a best case scenario to correctly assess the performances of the inference method for incomplete (in time and state space) data in further research. More generally, our study lays the foundations of the inference approach based on partially observed integrated diffusions that we are currently investigating (short discussion in Sect. 5).

2 Construction of the diffusion approximation of epidemic models

In this section we present the generic procedure for building the diffusion approximation of a density dependent jump Markov process $(Z(t))$ in $\mathbb{Z}^P$ as proposed in Ethier and Kurtz (2005). Then, we derive it for two $SIR$-like epidemic models and extend it to general density time-dependent Markov processes. A first normalization, corresponding to a law of large numbers, provides the convergence of $(Z(t))$ to a deterministic limit $x(t)$, solution of an ordinary differential equation. Then, centering $(Z(t))$, a central limit theorem yields that the process $\sqrt{N}(\frac{Z(t)}{N} - x(t))$ is approximated either by a Gaussian process (Van Kampen 1992) or by a diffusion process, these two approximations being essentially equivalent processes at least on fixed time intervals (see Ethier and Kurtz 2005, Chapter 11, Section 2.3).

In fact, the diffusion approximation possesses a small diffusion coefficient proportional to $N^{-1/2}$ and the Gaussian process comes from large deviations and corresponds to the first two terms of the diffusion expansion (see Azencott 1982; Freidlin and Wentzell 1978). We chose here the diffusion approximation, as our theoretical results supporting the estimation of epidemic model parameters are built on it (but they are still valid for the Gaussian approximation of the Markov jump process).

2.1 A different representation of Markov jump processes (Ethier and Kurtz 2005)

A multidimensional Markov jump process $(Z(t), t \geq 0)$ with state space $E \subset \mathbb{Z}^P$ is usually described by an initial distribution $\lambda(.)$ on $E$, and a collection of non negative functions $(\alpha_l(.) : E \to \mathbb{R}^+)$ indexed by $l \in \mathbb{Z}^P$, $l \neq (0, \ldots, 0)$ that satisfy,

$$\forall k \in E, \ 0 < \sum_{l \in \mathbb{Z}^P} \alpha_l(k) = \alpha(k) < \infty. \quad (1)$$
These functions represent the transition intensities of \((Z(t))\) by setting the transition rates from \(k\) to \(k + l\),

\[
q_{k,k+l} = \alpha_l(k).
\]  

The collection \((\alpha_l(k))_l\) is associated to all the possible jumps from state \(k\), and the time spent in this state is exponentially distributed with parameter \(\alpha(k)\). The transition probabilities of the embedded Markov chain \((C_m)\) are, for \(k\) and \(k + l \in E\),

\[
P(C_{m+1} = k + l | C_m = k) = \frac{\alpha_l(k)}{\alpha(k)}.
\]  

The generator \(A\) of the Markov jump process \((Z(t), t \geq 0)\) is defined on the set of real measurable and bounded functions \(f\) on \((E, \mathcal{B}(E))\) and writes,

\[
A f(k) = \sum_{l \in \mathbb{Z}^p} \alpha_l(k)(f(k + l) - f(k)) = \alpha(k) \sum_{l \in \mathbb{Z}^p} (f(k + l) - f(k)) \frac{\alpha_l(k)}{\alpha(k)}.
\]  

Following Chapter 6, Section 4 of Ethier and Kurtz (2005), there is another useful expression based on Poisson processes for \(Z(t)\). Let \((P_l(.))\) be a family of independent Poisson processes with rate 1, indexed by \(l \in \mathbb{Z}^p\), independent of \(Z(0)\). After applying to each \(P_l(.)\) a random time change based on \(\alpha_l(.)\), \(Z(t)\) can be expressed as,

\[
Z(t) = Z(0) + \sum_{l \in \mathbb{Z}^p} l \int_0^t \alpha_l(Z(u))du.
\]  

This new expression of the jump process \(Z(t)\), obtained by proving the equality of the two infinitesimal generators associated to each representation, is very powerful to evaluate distances between trajectories of processes, and consequently to establish approximation results. All technical details are provided in Appendix A.1.

**Remark 1** From the point of view of simulation algorithms for epidemic dynamics, the classical representation based on transition rates (2) relates to the algorithm of Gillespie (1977), whereas the time changed Poisson representation (4) corresponds to the more general scheme of Sellke (1983) (typically for non exponential distributions of the infectious period).

### 2.2 The generic approach for the diffusion approximation

Let us now consider density dependent Markov jump processes on \(\mathbb{Z}^P\). The Markov process \(Z(t)\) has state space \(E = \{0, \ldots, N\}^P\), where \(N\) (fixed parameter) represents the total population size, transition rates \(q_{k,k+l} = \alpha_l(k)\) and jumps \(l\) in \(E^- = \{-N, \ldots, N\}^P\) such that \(k, k + l \in E, l \neq \{0\}^P\).

For \(l \in \mathbb{Z}^P\) or \(y \in \mathbb{R}^P\) with components \(y_i\), denote by \(\vec{l}\) and \(\vec{y}\) the transpositions of \(l\) and \(y\) respectively, and by \([y]\) the vector of \(\mathbb{Z}^P\) with components \([y_i]\), where \([y_i]\) is the integer part of \(y_i\).
In order to describe the behavior of \((Z(t))\) for large \(N\), we assume,

\[(H1): \forall l \in E^-, \forall y \in [0, 1]^p \quad \frac{1}{N} \alpha_l([Ny]) \xrightarrow{N \to +\infty} \beta_l(y),\]

\[(H2): \forall l \in E^-, \quad \beta_l \in C^2([0, 1]^p).\]

These assumptions ensure that the Markov jump process is a density dependent process (H1) with asymptotic regularity properties for its transition rates (H2). Note that for density dependent processes, the collection of functions \(\alpha_l\) and \(\beta_l\) is finite and the condition \(\alpha(k) < \infty\) of (1) is satisfied; for more general processes, additional assumptions on \(\beta_l\) similar to (1) are required.

From the original jump process \((Z(t))\) on \(E = \{0, \ldots, N\}^p\), we consider the normalized jump Markov process \((Z_N(t) = \frac{Z(t)}{N})\) on \(E_N = \{N^{-1}k, \ k \in E\}\). It satisfies, using (2), \(Z_N(0) = \frac{Z(0)}{N}\) and for \(y, y + z \in E_N\),

\[q^N_{y, y+z} = \alpha_N(z) ; \quad A_N f(y) = \sum_{l \in E^-} \alpha_l(Ny) \left( f\left(y + \frac{l}{N}\right) - f(y)\right).\]

The time changed Poisson process representation (4) is

\[Z_N(t) = Z_N(0) + \sum_{l \in E^-} \frac{l}{N} P_l \left( \int_0^t \alpha_l(NZ_N(u)) du\right).\]

In order to build approximation processes from \((Z_N(t))\), the first step is to assess the mean behavior of \(Z_N(t)\) as \(N \to +\infty\) (which yields a sort of “law of large numbers” for \(Z_N\)) and the second step consists in specifying what happens around the mean (“Central Limit Theorem”). Heuristically, this can be obtained either by expanding in Taylor series the generator \(A_N\) of \(Z_N\) or by studying the paths \((Z_N(t))\) by means of expression (6). We sketch below the two perspectives for the behavior both at the mean and around it.

First, to study \(Z_N(t)\) at its mean, let us define the two functions \(b_N\) and \(b\) on \([0, 1]^p\) by

\[b_N(y) = \sum_{l \in E^-} \frac{l}{N} \alpha_l(Ny) \quad \text{and} \quad b(y) = \sum_{l \in E^-} l\beta_l(y).\]

Under (H1), \(b_N(.)\) converges to \(b(.)\) uniformly on \([0, 1]^p\).

Using definitions (5) and (7), the generator \(A_N\) of \(Z_N\) writes, for \(y \in E_N\), \(f(.) \in C^1(\mathbb{R}^p)\) with gradient \(\nabla f(y)\) and \(l = (l_1, \ldots, l_p)\),

\[A_N f(y) = \sum_{l \in E^-} \alpha_l(Ny) \left( f\left(y + \frac{l}{N}\right) - f(y) - \frac{1}{N} \sum_{i=1,\ldots,p} l_i \frac{\partial f}{\partial y_i}(y)\right) + b_N(y) \cdot \nabla f(y).\]
Under (H1), the first term of $A_N f(y)$ goes to 0, and the second term converges using (7) to $b(y) \nabla f(y)$. Therefore, as $N \to \infty$, $(Z_N(t))$ converges in distribution to the process with generator $A f(y) = b(y) \nabla f(y)$. The function $b$ inherits the regularity properties of $\beta_l$, so $b$ is Lipschitz by (H2) and the ODE $\frac{dx}{dt} = b(x(t))$ has a unique well defined regular solution $x(t)$ satisfying

$$x(t) = x_0 + \int_0^t b(x(u))du. \quad (8)$$

Besides, a stronger result holds for $(Z_N(.))$ (i.e. “law of large numbers”). If $Z_N(0) \xrightarrow{N \to +\infty} x_0$, then, under (H1)–(H2),

$$\forall t \geq 0, \lim_{N \to \infty} \sup_{u \leq t} \| Z_N(u) - x(u) \| = 0 \text{ a.s.} \quad (9)$$

Second, to specify the asymptotic behavior of the process $(Z_N(t))$ around its deterministic limit $x(t)$, we have to pursue our approach for the mean by either expanding further $A_N$, or by setting a “Central Limit Theorem” for $(Z_N(t))$. The two approaches lead to different approximations. Indeed, the first one leads to a diffusion process $X_N(t)$, whereas the second approach developed in Chapter 11, Section 2 of Ethier and Kurtz (2005) consists in studying the process $\sqrt{N}(Z_N(t) - x(t))$ based on the expression (6) of $Z_N(.)$ in the specific case of transition rates $\alpha_l(.)$ such that

$$\forall l \in E^-, \forall y \in [0,1]^p, \forall N, \frac{1}{N} \alpha_l([Ny]) = \beta_l(y). \quad (10)$$

Let us define the two $p \times p$ positive symmetric matrices $\Sigma_N$ and $\Sigma$ from the $(\alpha_l)_{l \in E^-}$ and $(\beta_l)_{l \in E^-}$ families.

$$\Sigma_N(y) = \frac{1}{N} \sum_{l \in E^-} \alpha_l(y) l \ l' \text{ and } \Sigma(y) = \sum_{l \in E^-} \beta_l(y) l \ l'. \quad (11)$$

Expanding in Taylor series $A_N f(y)$ yields, using (7) and (11), for $f \in C^2(\mathbb{R}^p, \mathbb{R})$,

$$A_N f(y) = \sum_{l \in E^-} \alpha_l(Ny) \left( f\left(y + \frac{l}{N}\right) - f(y) - \frac{1}{N} \sum_{i=1}^d l_i \frac{\partial f}{\partial y_i}(y) - \frac{1}{2N^2} \sum_{i,j=1}^p l_il_j \frac{\partial^2 f}{\partial y_i \partial y_j}(y) \right)$$

$$+ b(y) \nabla f(y) + \frac{1}{2N} \sum_{i,j=1}^p \Sigma_{ij}(y) \frac{\partial^2 f}{\partial y_i \partial y_j}(y)$$

$$+ (b_N(y) - b(y)) \nabla f(y) + \frac{1}{2N} \left( \sum_{i,j=1}^p (\Sigma_N - \Sigma)_{ij}(y) \frac{\partial^2 f}{\partial y_i \partial y_j}(y) \right).$$
Heuristically, the first term of $A_N f(y)$ is of $O(1/N^2)$, the second term corresponds to the ODE; the second and third terms correspond to the generator of a $p$-dimensional diffusion process $(X_N(t))$ with drift function $b(y)$ defined in (7) and diffusion matrix $\Sigma(y)$ defined in (11). Under (H1), the last term of $A_N f(y)$ is of order $o(1/N)$. An additional assumption on the $a_l$ is required to ensure that the remaining term $(b_N(y) - b(y)) \nabla f(y)$ is also $o(1/N)$. Let us define more precisely the diffusion $X_N$. Let $(B(t)_{t \geq 0})$ be a $p$-dimensional standard Brownian motion on a probability space $\mathbb{P} = (\Omega, (\mathcal{F}_t)_{t \geq 0}, P)$. Assume that $Z_N(0)$ is $\mathcal{F}_0$-measurable, then $X_N$ will be the solution of the stochastic differential equation,

$$dX_N(t) = b(X_N(t)) \, dt + \frac{1}{\sqrt{N}} \sigma(X_N(t)) \, dB(t); \quad X_N(0) = Z_N(0),$$

where $\sigma(y)$ is a $p \times p$ matrix satisfying $\sigma(y)^t \sigma(y) = \Sigma(y)$.

This is a Markov process with generator $B_N$ such that, for $f \in C^2(\mathbb{R}^p)$,

$$B_N f(y) = \frac{1}{2N} \sum_{i,j=1}^p \Sigma_{ij}(y) \frac{\partial^2 f}{\partial y_i \partial y_j}(y) + b(y) \cdot \nabla f(y).$$

The two generators $A_N$ and $B_N$ satisfy $\| A_N f - B_N f \| = o(1/N)$, which suggests the approximation of $(Z_N(.))$ by $(X_N(.))$.

Let us now briefly recall the results of Ethier and Kurtz (2005) (Theorem 1, Chapter 11, Section 3) concerning $Y_N(t) = \sqrt{N}(Z_N(t) - x(t))$. For this, define $\Phi(t, u)$, the resolvent matrix of the “linearized” ODE satisfied by $x(.)$, where $\nabla b(y)$ denotes the $p \times p$ matrix $(\frac{\partial b_i}{\partial y_j}(y))$:

$$\frac{d\Phi}{dt}(t, u) = \nabla b(x(t)) \Phi(t, u); \quad \Phi(u, u) = I_p.$$

Then, for transition rates satisfying (10), Ethier and Kurtz (2005) use the Poisson decomposition (6) to get that $(Y_N(t))$ converges in distribution to a centered Gaussian process $(G(t))$, with covariance matrix,

$$\text{Cov}(G(t), G(r)) = \int_0^{t \wedge r} \Phi(t, u) \Sigma(x(u))^t \Phi(r, u) du.$$

The rigorous proof that $(Y_N(t))$ converges in distribution to $(G(t))$ in the case where the $(a_l(\.))_l$ just satisfy (H1) and (H2) is surprisingly difficult. Moreover, the Poisson decomposition (6) can no longer be used for the time-dependent processes that we consider in the next section. Hence, we prove the diffusion approximation for non homogeneous systems using convergence theorems for semimartingales under the assumption:
(H1)′: \( \forall l \in E^-, \sup_{y \in [0,1]} \| \sqrt{N} \left( \frac{1}{N} \alpha_l([N y]) - \beta_l(y) \right) \| \rightarrow 0 \).

Some details along with a generalization are given in Appendix A.2. Note that condition (H1)′ does not seem to ensure that \( \| A_N - B_N \| = o(1/N) \) (where \( \sqrt{N} \) has to be replaced by \( N \)). But the proof given in the appendix, which is more precise, requires only (H1)′.

Let us now set the different links between these two limit processes. Although the “Central Limit Theorem” and the generator expansion approach result in two different limiting processes, the theory of random pertubations of dynamical systems (or the stochastic Taylor expansion of the diffusion, Freidlin and Wentzell 1978; Azencott 1982) clarifies the link between the Gaussian process \( G(t) \) defined in (14) and the diffusion process \( X_N(t) \) defined in (12). Indeed, setting \( \epsilon = 1/\sqrt{N} \), the paths \( X_N(.) \) satisfy,

\[ X_N(t) = X_\epsilon(t) = z(t) + \epsilon g(t) + \epsilon R_\epsilon(t), \]

where \( \sup_{t \leq T} \| \epsilon R_\epsilon(t) \| \rightarrow 0 \) in probability as \( \epsilon \rightarrow 0 \), and where \( z(t) \), \( g(t) \) are defined as follows. The function \( z(t) \) satisfies the ODE \( \frac{dz}{dt} = b(z(t)) dt \); \( z(0) = x_0 \). Therefore this is precisely the previous solution \( x(t) \). The process \( g(t) \) satisfies the stochastic differential equation

\[ dg(t) = \nabla b(x(t))g(t) dt + \sigma(x(t))dB(t) \quad ; \quad g(0) = 0. \]

This SDE can be solved explicitly, and its solution is the process

\[ g(t) = \int_0^t \Phi(t, s)\sigma(x(s))dB(s). \]

Hence, \( g(t) \) is a centered Gaussian process having the same covariance matrix as the Gaussian process \( G \) defined in (14).

For statistical purposes, this result is very useful. Indeed, it was the starting point of the results in Guy et al (2014).

We can summarize the approximation of our epidemic diffusion model for statistical purposes in the following algorithm. From now on, \( (\alpha_l) \) and the derived functions will depend on parameters \( (\theta) \).

Step 1. Write all the mechanistic transitions between states and their respective intensities (functions \( \alpha_l \)).

Step 2. Compute associated functions \( \beta_l \) derived from (H1).

Step 3. Compute functions \( b(\theta, y) \) and \( \Sigma(\theta, y) \) (defined in (7) and (11) respectively) from \( \beta_l \).

2.3 Building the diffusion approximation for the SIR epidemic model

We consider the simple SIR model defined in Sect. 1 through the bidimensional continuous-time Markov jump process \( Z_N(t) \). Following the three-step algorithm introduced above, we build the associated SIR diffusion process.
Step 1. The process $Z_N(t)$ has the state space $\{0, \ldots, N\}^2$ and the jumps $l$ are $(-1, 1)$ and $(0, -1)$. The transition rates are respectively
\[ q((s, i), (s-1, i+1)) = \lambda s \frac{1}{N} = \alpha_{(-1,1)}(S, I) \quad \text{and} \quad q((s, i), (s, i-1)) = \gamma I = \alpha_{(0,-1)}(S, I). \]

Step 2. Let $y = (s, i) \in [0, 1]^2$ and the parameter $\theta = (\lambda, \gamma)$. Then,
\[ \frac{1}{N} \alpha_{(-1,1)}([N y]) = \frac{1}{N} \lambda s [N i] \xrightarrow{N \to +\infty} \beta_{(-1,1)}(s, i) = \lambda s i \quad ; \]
\[ \frac{1}{N} \alpha_{(0,-1)}([N y]) = \frac{1}{N} \gamma [N i] \xrightarrow{N \to +\infty} \beta_{(0,-1)}(s, i) = \gamma i. \]

(H1)–(H2) are satisfied.

Step 3. Function $b(\theta, y)$ defined in (7) is then
\[ b((\lambda, \gamma), (s, i)) = \lambda s i \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \gamma i \begin{pmatrix} 0 \\ -1 \end{pmatrix} = \begin{pmatrix} -\lambda s i \\ \lambda s i - \gamma i \end{pmatrix}. \]

The diffusion matrix $\Sigma(\theta, y)$ defined in (11) writes as
\[ \Sigma((\lambda, \gamma), (s, i)) = \lambda s i \begin{pmatrix} -1 \\ 1 \end{pmatrix} (1 1) + \gamma i \begin{pmatrix} 0 \\ -1 \end{pmatrix} (0 1) = \begin{pmatrix} \lambda s i - \lambda s i \\ -\lambda s i \lambda s i + \gamma i \end{pmatrix}. \]

Assume that $(\frac{\xi_0}{N}, \frac{\eta}{N}) \xrightarrow{N \to +\infty} (s_0, i_0)$, and let $\sigma(\theta, y)$ be a square root of $\Sigma(\theta, y)$ and $(B(t) = \frac{1}{4}(B_1(t), B_2(t)))$ denote a standard two-dimensional Brownian motion.

Then, setting $\sigma(\theta, (s, i)) = \begin{pmatrix} \sqrt{\lambda s i} & 0 \\ -\sqrt{\lambda s i} & \sqrt{\gamma i} \end{pmatrix}$ yields that $X_N(t) = \begin{pmatrix} S_N(t) \\ I_N(t) \end{pmatrix}$ satisfies $X_N(0) = x_0$ and
\[
\begin{align*}
    dS_N(t) &= -\lambda S_N(t) I_N(t) dt + \frac{1}{\sqrt{N}} \sqrt{\lambda S_N(t) I_N(t)} dB_1(t) \\
    dI_N(t) &= (\lambda S_N(t) I_N(t) - \gamma I_N(t)) dt - \frac{1}{\sqrt{N}} \sqrt{\lambda S_N(t) I_N(t)} dB_1(t) - \sqrt{\gamma I_N(t)} dB_2(t).
\end{align*}
\]

2.4 The diffusion approximation for the non autonomous case: the SIRS model with seasonal forcing

While the SIR model is suited for studying a single outbreak, it is not appropriate for describing recurrent epidemics. In order to reproduce a series of outbreaks, we need to compensate the depletion of susceptibles by other mechanisms, such as demography (with birth and death rates equal to $\mu$ for a stable population of size $N$) and/or reinsertion of removed individuals into $S$ compartment (as a consequence of immunity waning, after, on average, $1/\delta$ time). This leads to the SIRS model. We also add a new term to the transition $S \to I$ that writes now as $(S, I) \xrightarrow{\delta} (S-1, I+1)$. This modification is introduced in order to avoid extinction, more likely to occur when simulating recurrent epidemics based on Markov jump process. The new term can be interpreted as constant immigration flow at rate $\eta$ in the infected class.

The diffusion approximation of this model obtained according to the scheme introduced in Sect. 2.2 is:

**Steps 1 & 2.** $(S, I) \xrightarrow{\delta} (S-1, I+1) \Rightarrow \beta_{(-1,1)}(s, i) = \lambda s (i + \eta),$

$(S, I) \xrightarrow{(\gamma + \mu) I} (S, I-1) \Rightarrow \beta_{(0,-1)}(s, i) = (\gamma + \mu) i.$
\[(S, I) \xrightarrow{\mu_S} (S-1, I) \Rightarrow \beta_{(-1,0)}(s, i) = \mu s \text{ and} \]
\[(S, I) \xrightarrow{\mu N + \delta (N - S - I)} (S + 1, I) \Rightarrow \beta_{(1,0)}(s, i) = \mu + \delta (1 - s - i). \]

We can notice that in the SIRS population dynamics with multiple epidemic waves, the proportion of infected individuals (the signal) is generally low \((\sim 10^{-3})\). Consequently, in order to obtain a satisfying ratio (greater than 1) between signal and noise, it is necessary to consider large populations \((N \sim 10^6)\), since the noise has an order of \(1/\sqrt{N}\).

Although able to describe more than one epidemic wave, it is well known that the SIRS model leads to a function \(b(y)\) and its associated ODE solution \((s(t), i(t))\), for which oscillations vanish (Keeling and Rohani 2011, Chapter 5) as \(t \to \infty\) (and so does the diffusion, Fig. 1).

To overcome this problem, a natural assumption to be considered is that the transmission is seasonal. Mathematically, this leads to a time non homogeneous transmission rate, expressed under a periodic form

\[
\lambda(t) := \lambda_0 (1 + \lambda_1 \sin(2\pi t/T_{per}))
\]

where \(\lambda_0\) is the baseline transition rate, \(\lambda_1\) the intensity of the seasonal effect on transmission and \(T_{per}\) the period of the seasonal trend (see Keeling and Rohani 2011, Chapter 5).

These transition rates define a time-inhomogeneous Markov jump process, for which the martingale problem is solved by means of an extended generator (Ethier and Kurtz 2005, Chapter 4, Section 7). However, these authors studied diffusion and Gaussian approximations of Markov jump processes only in the case of time homogeneous transition rates.

---

**Fig. 1** Proportion of infected individuals over time for the diffusion approximation (blue) and the corresponding ODEs (black) of the SIRS model with \(N = 10^7\), \(T_{per} = 365\), \(\mu = 1/(50 \times T_{per})\), \(\eta = 10^{-6}\), \((s_0, i_0) = (0.7, 10^{-4})\), and \((\lambda_0, \gamma, \delta) = (0.5, 1/3, 1/(2 \times 365))\), without seasonality, \(\lambda_1 = 0\) (a) and with seasonality, \(\lambda_1 = 0.02\) (b)
Using that density dependent Markov processes can be viewed as semimartingales with specific characteristics, we studied the limit processes with another approach, based on convergence theorems for semimartingales (Jacod and Shiryaev 1987). (see Appendix A.2 for details).

Assumptions (H1), (H2), (H1)′ are modified as follows:

\( (H1t) \): \( \forall (l, t, y) \in E^- \times [0, T] \times [0, 1]^p, \frac{1}{N} \alpha_l(t, [Ny]) \rightarrow_{N \rightarrow +\infty} \beta_l(t, y) \),

\( (H2t) \): \( \forall l \in E^-, \beta_l(. . .) \in C^2([0, T] \times [0, 1]^p) \).

\( (H1l) : \forall l \in E^-, \sup_{y \in [0, 1], t \in [0, T]} \| \sqrt{N} (\frac{1}{N} \alpha_l(t, [Ny]) - \beta_l(t, y)) \| \rightarrow_{N \rightarrow +\infty} 0. \)

The new system obtained is still of dimension 2 (assuming a constant population size) with four transitions for the corresponding Markov jump process. The procedure introduced in Sect. 2.2 can be generalized to time dependent models:

**Step 1.** \( (S, I) \xrightarrow{\mu S} (S - 1, I + 1) \),

\( (S, I) \xrightarrow{(y + \mu) I} (S, I - 1) \), and

\( (S, I) \xrightarrow{\mu N + \delta(N - S - 1)} (S + 1, I) \).

**Step 2.** The rate of the first transition writes as

\[ \frac{1}{N} \alpha(-1,1)(t, S, I) = \lambda(t) \frac{S}{N} \xrightarrow{N \rightarrow +\infty} \beta(-1,1)(t, (s, i)) = \lambda(t)s(i + \eta). \]

Since the time dependence satisfies

\[ \alpha_l(t, k) = \lambda(t) \alpha_l'(k), \]

and since only a space normalization is applied, the generic results from Sect. 2.2 still hold by replacing functions \( \beta_l(y) \) with \( \beta_l(t, y) \). The three other transitions are identical to those of the SIRS model without seasonality.

**Step 3.** Hence, for \( \theta = (\lambda_0, \lambda_1, \gamma, \delta, \eta, \mu) \), we obtain the drift term and diffusion matrix as:

\[ b(t, \theta, y) = \lambda(t)s(i + \eta) \begin{pmatrix} -1 & 1 \end{pmatrix} + (\gamma + \mu)i \begin{pmatrix} 0 & 1 \end{pmatrix} + (\mu + \delta(1 - s - i)) \begin{pmatrix} 0 & 1 \end{pmatrix} \]

\[ + \mu s \begin{pmatrix} -1 & 0 \end{pmatrix} \]

\[ = \begin{pmatrix} -\lambda(t)s(i + \eta) + \delta(1 - s - i) + \mu(1 - s) \lambda(t)s(i + \eta) - (\gamma + \mu)i \end{pmatrix} = \begin{pmatrix} b_1(t, \theta, y) \\ b_2(t, \theta, y) \end{pmatrix}, \]

\[ \Sigma(t, \theta, y) = \lambda(t)s(i + \eta) \begin{pmatrix} -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 \end{pmatrix} + (\gamma + \mu)i \begin{pmatrix} 0 & 1 \end{pmatrix} + \mu s \begin{pmatrix} -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \end{pmatrix} \]

\[ = \begin{pmatrix} \lambda(t)s(i + \eta) + \delta(1 - s - i) + \mu(1 + s) - \lambda(t)s(i + \eta) \lambda(t)s(i + \eta) + (\gamma + \mu)i \end{pmatrix}. \]

The Cholesky decomposition yields for \( y = (s, i) \),

\[ \sigma(t, \theta, y) = \begin{pmatrix} \sigma_{1,1}(t, \theta, y) & 0 \\ \sigma_{2,1}(t, \theta, y) & \sigma_{2,2}(t, \theta, y) \end{pmatrix} \]

with
\[\begin{align*}
\sigma_{1,1}(t, \theta, y) &= \sqrt{\lambda(t) s(i + \eta)} + \delta(1 - s - i) + \mu(1 + s), \\
\sigma_{1,2}(t, \theta, y) &= -\sqrt{\frac{\lambda(t) s(i + \eta)}{\lambda(t) s(i + \eta) + \delta(1 - s - i) + \mu(1 + s)}} \quad \text{and} \\
\sigma_{2,2}(t, \theta, y) &= \sqrt{\frac{\lambda(t)(y + \mu) s(i + \eta) + \frac{\lambda(t) s(i + \eta)}{\lambda(t) s(i + \eta) + \delta(1 - s - i) + \mu(1 + s)}(1 - s - i) + \mu(1 + s) + (y + \mu) i (1 - s - i) + \mu(1 + s))}.
\end{align*}\]

Finally, we get the diffusion \(X_N(t) = (S_N(t), I_N(t))\) starting from \(X_N(0) = (s_0, i_0)\),

\[
\begin{cases}
    dS_N(t) = b_1(t, \theta; S_N(t), I_N(t)) + \frac{1}{\sqrt{N}} \sigma_{1,1}(t, \theta; S_N(t), I_N(t)) dB_1(t) \\
    dI_N(t) = b_2(t, \theta; S_N(t), I_N(t)) + \frac{1}{\sqrt{N}} \sigma_{2,1}(t, \theta; S_N(t), I_N(t)) dB_1(t) + \frac{1}{\sqrt{N}} \sigma_{2,2}(t, \theta; S_N(t), I_N(t)) dB_2(t).
\end{cases}
\]

### 3 Minimum contrast estimators for diffusion processes

The statistical inference for continuously observed diffusion processes on a finite interval is based on the likelihood of the diffusion and obtained using the Girsanov formula (see e.g. Lipster and Shiryaev 2001 for the asymptotics \(T \to \infty\) and Kutoyants 1984 in the asymptotics of small diffusion coefficient \(\epsilon \to 0\)). Discretely observed diffusion processes are discrete time Markov processes and thus their likelihood depends on the transition densities of the diffusion \(p_\theta(t_k - 1, t_k; x, dy) = \mathbb{P}_\theta(X(t_k) \in [y, y + dy] | X(t_{k-1}) = x)\). Since the dependence with respect to the parameters \(\theta\) of these transition densities is not explicit, the likelihood is untractable and other approaches have been proposed. This situation often occurs for stochastic processes, and other processes than the likelihood can be used to estimate parameters. These processes can be good approximations of the likelihood or can be completely different. For independent random variables, the estimators obtained with such approaches are called \(M\)-estimators (see van der Vaart 2000). For stochastic processes, the processes used instead of the likelihood are often called contrast processes with associated minimum contrast estimators. They have to satisfy a series of conditions to lead to good estimators. Contrary to the i.i.d. parametric set-up, there is no well recognized terminology to name these estimators, and we have adopted here the terminology of contrast processes and minimum contrast estimators.

In a previous work (Guy et al 2014) we developed a parametric inference approach for discretely observed multidimensional diffusions with small diffusion coefficient \(\epsilon = 1/\sqrt{N}\) (for \(N\) large). The diffusion is observed on interval \([0, T]\) at times \(t_k = k\Delta\), for \(k = 1, \ldots, n\) \((T = n\Delta)\). We provided minimum contrast estimators with good properties: consistent and asymptotically normal for both drift and diffusion parameters for small sampling interval and for drift parameters in the case of fixed sampling interval. Let us stress that, for general diffusions with small diffusion coefficient observed on a fixed time interval, two different asymptotics can be considered. The first one corresponds to the small diffusion asymptotics \((\epsilon = 1/\sqrt{N} \to 0 \Leftrightarrow N \to +\infty)\) and the second one corresponds to the sampling interval going to zero \((\Delta = \Delta_n \to 0 \Leftrightarrow n \to +\infty)\). When the two asymptotics occur simultaneously, the rates of convergence of parameters in the drift and diffusion coefficient differ: drift parameters at rate \(\epsilon^{-1}\) and diffusion parameters at rate \(\sqrt{n}\). Consequently, for small sampling interval and
for diffusion approximations where the same parameter $\theta$ is present in the drift and diffusion coefficients simultaneously, we can choose the most efficient rate to estimate this parameter. Here, we introduce a new variant of the contrast of Guy et al (2014) (Section 3.3.1). This contrast is developed for the special case where the parameters of drift and diffusion terms are identical in the asymptotics $N \to +\infty$ and for $n$ fixed. All these characteristics fit well the epidemic framework (e.g. large population size and limited number of observations). The new contrast improves the non asymptotic accuracy of related estimators while preserving their asymptotic properties.

3.1 The main lines of the inference method

From now on, we assume that the parameter set $\Theta$ is a compact subset of $\mathbb{R}^m$, and that the true value of the parameter $\theta_0$ belongs to $\Theta$.

As stated in previous sections, only the computation of functions $b$ and $\Sigma$ is required to build the approximation diffusion (12) of the Markov jump process. These two functions allow building a family of contrast processes for discrete observations at times $t_k$, $k \in \{0, \ldots, n\}$.

Using (7),(8),(13) leads to $x_\theta(t)$, $b(\theta, x(t))$, the resolvent matrix $\Phi_\theta$, and the Gaussian process $g_\theta(t)$ as the limit of $\sqrt{N}(Z_N(t) - x_\theta(t))$. Then, we can state the fundamental property of our contrast approach. The Gaussian process $g_\theta$ satisfies:

$$g_\theta(t_k) = \Phi_\theta(t_k, t_{k-1}) g_\theta(t_{k-1}) + \sqrt{\Delta} V_k^\theta,$$

with $(V_k^\theta)_k$ a sequence of $n$ independent centered Gaussian vectors with bounded covariance matrix, and $\Delta = T/n$ the sampling interval. The sequence $(V_k^\theta)$ being independent, we can compute its likelihood and derive a contrast process for the diffusion.

For this, let us define the function $A_k(\theta, (X_{tk})_{k \in \{0, \ldots, n\}}) = A_k(\theta)$ for the diffusion $(X_N(t))$ at time points $(t_k)_{k \in \{0, \ldots, n\}}$,

$$A_k(\theta) = X_{tk} - x_\theta(t_k) - \Phi_\theta(t_k, t_{k-1}) [X_{tk-1} - x_\theta(t_{k-1})].$$

Let us also introduce the matrix $S_k^\theta$, corresponding to the covariance matrix of the family $(V_k^\theta)$ as

$$S_k^\theta = \frac{1}{\Delta} \int_{t_{k-1}}^{t_k} \Phi_\theta(t_k, u) \Sigma(\theta, x_\theta(u)) \Sigma(\theta, x_\theta(u))^t \Phi_\theta(t_k, u) du.$$

This leads to the construction of the contrast process $U_N$ and the associated estimator $\hat{\theta}_N$:

$$U_N(\theta, (X_{tk})_{k \in \{0, \ldots, n\}}) = \sum_{k=1}^n \left[ \frac{1}{N} \log \left( \det \left( S_k^\theta \right) \right) + \frac{1}{\Delta} A_k(\theta) (S_k^\theta)^{-1} A_k(\theta) \right],$$

$$\hat{\theta}_N = \arg\min_{\theta \in \Theta} U_N(\theta, (X_{tk})_{k \in \{0, \ldots, n\}}).$$
In this contrast process, $A_k(θ)$ can be interpreted as an error function between observations and the deterministic trajectory associated to the parameter $θ$ at time $t_k$, incorporating the propagation of the error at time $t_{k-1}$, and $S^θ_k$ as a corrective weight matrix. The contrast (20) is a modified version of a contrast proposed in our previous work (Section 3.3.1 of Guy et al. 2014). The main improvement is provided by the additional term $1/N \log(\det(S^θ_k))$ in (20) which corrects a non asymptotic bias of $\hat{θ}_N$ (noticed in simulations presented in Guy et al. 2014), while preserving its asymptotic properties as $N \to \infty$ and $n$ fixed. Since $\sum_{k=1}^{n} 1/N \log(\det(S^θ_k))$ is a finite sum of bounded terms, it will tend to 0 as $N \to \infty$. According to Proposition 3.2 in Guy et al. (2014), $\hat{θ}_N$ is consistent and asymptotically normal:

$$\sqrt{N}(\hat{θ}_N - θ_0) \xrightarrow{L} N_m(0, I^{-1}(n, θ_0))$$

where

$$I(n, θ_0) = \left( \sum_{k=1}^{n} D_{k,i} \left( S^θ_k \right)^{-1} \right)_{1 \leq i, j \leq m}$$

with $D_{k,i} = -\frac{\partial x_θ(t_k)}{\partial θ_i}(θ_0) + \Phi_θ(t_k, t_{k-1}) \frac{\partial x_θ(t_{k-1})}{\partial θ_i}(θ_0)$.

It is important to point out that the above results are still valid for any number of observations $n$. As $n$ increases, the asymptotic information $I(n, θ_0)$ increases (and consequently the width of confidence intervals decreases) towards the efficient bound corresponding to the continuous observation of the diffusion on $[0, T]$ for parameters in the drift functions:

$$I(n, θ_0) \xrightarrow{n \to \infty} I_b(θ_0)$$

$$= \left( \frac{1}{T} \int_0^T \frac{\partial b(θ, x_θ(t))}{\partial θ_i}(θ_0) \Sigma^{-1}(θ_0) \frac{\partial b(θ, x_θ(t))}{\partial θ_j}(θ_0) dt \right)_{1 \leq i, j \leq m}.$$

**Remark 2** For irregular sampling interval, $\hat{θ}_N$ will still keep its properties (see Appendix A.3 for more details). This aspect has practical implications since it can be used in various observed designs of epidemics: for instance, many data points could be recorded in the early phase of the epidemic and much less in the second phase.

### 3.2 Case of time dependence

As stated in Sect. 2.4, the diffusion approximation holds for time non homogeneous Markov jump processes. This leads to drift and diffusion functions $b$ and $Σ$ which are time dependent. Although the results obtained in Guy et al. (2014) were proved only for autonomous diffusions $dX_N(t) = b(θ, X_N(t)) dt + \frac{1}{\sqrt{N}} Σ(θ, X_N(t)) dB_t$, $x_0 \in \mathbb{R}^p$, they can extend to time dependent diffusion processes. Previous quantities need to be modified by replacing each occurrence of $b(θ, x_θ(t))$, $Σ(θ, x_θ(t))$ and $σ(θ, x_θ(t))$ by $b(t, θ, x_θ(t))$, $Σ(t, θ, x_θ(t))$ and $σ(t, θ, x_θ(t))$. The estimates of SIRS model parameters are obtained using this new framework (Sect. 4.2). Additional technical details are provided in Appendix A.3.
The inference method proposed in this study is assessed on simulated data. Two different epidemic models, the SIR and the SIRS with time-dependent transmission rate and demography, described in Sects. 2.3 and 2.4 respectively, are considered. Simulations are based on the algorithm of Gillespie (1977) for the SIR model and on the τ -leap method (Cao et al. 2005), more efficient for large populations, for the SIRS model. The accuracy of our minimum contrast estimators is investigated with respect to the population size \( N \), the number of observations \( n \), the parameter values and the model generating the data (Markov jump process and diffusion process). Only non extinct trajectories are considered for inference. An ad-hoc criterion (final epidemic size larger than 5% of the number of initial susceptibles) is chosen to define non extinction. For each set of parameter values, point contrast estimates (CE), theoretical confidence intervals (CI\(_{th}\)) and empirical ones (CI\(_{emp}\), built on 1,000 runs) are provided. Moreover, the intrinsic limits of the method are investigated by comparing CI\(_{th}\) for different values of \( n \), other parameters being fixed, with the theoretical variance co-variance matrix when \( n \to \infty \).

The first finding is that no relevant bias can be imputed to the model underlying the simulations, when comparing CE\(_s\) on data simulated under Markov jump and diffusion processes. Therefore, all the subsequent investigations were performed based on simulated trajectories with the Markov jump process. To facilitate the visual comparison of theoretical ellipsoids, they are all centered on the true parameter values for each scenario. Additionally, as CI\(_{emp}\) are very tight around point estimators, they are not represented on figures.

### 4.1 The SIR model

From now on, we consider the parameters of interest for epidemics: the basic reproduction number, \( R_0 = \frac{\lambda}{\gamma} \), which represents the average number of secondary cases generated by one infectious in a completely susceptible population, and the average infectious duration, \( d = \frac{1}{\gamma} \). The performances of our CE\(_s\) were evaluated for several parameter values (Table 1). For each combination of parameters, the analytical maximum likelihood estimator (MLE), calculated from the observation of all the jumps of the Markov process, was taken as reference.

As a general remark, we can say that the magnitude of the stochasticity of the sample path of \( I_N(t) \) depends on the value of \( R_0 \): for small \( R_0 \) the proportion of infected individuals in the population is smaller and so is the ratio signal over noise. Figure 2 illustrates the accuracy of the CE\(_s\) for a moderate population size \( N = 1,000 \) and from trajectories with weak \( (R_0 = 5) \) and strong \( (R_0 = 1.5) \) stochasticity. First, we can notice that there is a non negligible correlation between parameters \( R_0 \) and \( d \) (ellipsoids are deviated with respect to the \( Ox \) and \( Oy \) axes), increasing with \( d \) and decreasing with \( R_0 \), even for the MLE. Second, the shape of confidence ellipsoids (and consequently the projection on \( Ox \) and \( Oy \) axes with the largest CI\(_{th}\) among \( R_0 \) and \( d \)) depends on parameter values: e.g. CI\(_{th}\) is larger for \( R_0 \) than for \( d \) when \( R_0 = 5 \), whereas the opposite occurs for \( R_0 = 1.5 \). Third, for \( R_0 = 5 \) (i.e. for trajectories with...
Table 1  Range of parameters for the SIR model defined in Sect. 2.3

| Parameter | Description               | Values          |
|-----------|---------------------------|-----------------|
| $R_0$     | Basic reproduction number | 1.5, 3          |
| $d$       | Infectious period         | 3, 7 days       |
| $T^a$     | Final time of observation | 20, 40, 45, 100 days |
| $N$       | Population size           | 400, 1,000, 10,000 |
| $n$       | Number of observations    | 5, 10, 20, 40, 45, 100 |

$^a$ $T$ is chosen as the time point where the corresponding deterministic trajectory passes below the threshold of 1/100.

Fig. 2  Point estimators (plus symbol) built as averages over 1,000 independent simulated trajectories and their associated theoretical confidence ellipsoids for the SIR model: MLE with complete observations (red), CE for 1 obs/day, $n = 40$ (blue) and CE for $n = 10$ (black). Four scenarios are illustrated: $(R_0, d, T) = \{(1.5, 3, 40); (1.5, 7, 100); (5, 3, 20); (5, 7, 45)\}$, with $N = 1,000$. True parameter values are located at the intersection of horizontal and vertical dotted lines.

weak stochasticity), all the $CI_{th}$ are very close (especially those of MLE and CE for 1 obs/day), suggesting that there is no loss in estimation accuracy as expected for smooth trajectories, even when not all jumps are observed. This does not stand for $R_0 = 1.5$ when trajectories are very noisy (Fig. 3): the shape of ellipsoids and their relative positions vary with $n$. More specifically, for these trajectories, a large number of discretized observations ($n = 2,000$ which corresponds to the maximum number of possible jumps for $N = 1,000$ SIR dynamic with two types of transitions) does not compensate the loss of information compared to the case where all dates of jumps are observed. Finally, point values for MLE and CE calculated for different $n$ are very similar, which recommends the use of our CEs when only a small number of observations are available.
Figure 4 shows that the width of confidence intervals, when projecting on \(Ox\) and \(Oy\) axes, decreases with respect to \(\sqrt{N}\), whereas the correlation is not impacted by \(N\). For a given value of population size \(N\), confidence ellipsoids are relatively close except for the case of very few observations (e.g. \(n = 5\)). For the case \(N = 400\), the MLE is biased, mostly due to the strong variability in the trajectories. An explanation of this behavior lies in the fact that the MLE is optimal when data represent a “typical” realization of the Markov process, but could exhibit a bias when observations are far from the mean. This does not seem to occur when using our CEs. Although our method was introduced for large populations, it proves to be quite robust w.r.t. highly variable sample paths (obtained either for small \(N\) or small \(R_0\)).

We can see on Fig. 5 that even if on a large number of trajectories the asymptotic properties of MLE and CEs are very similar, for a particular trajectory, the estimation accuracy may be different. Beyond the intrinsic variability of point estimates, this can also be viewed as a consequence of the form of functions \(A_k(\theta)\). Indeed, they are more sensitive to variation in slope of the difference between deterministic trajectory and data than the classical least square distance.

**Remark 3** The robustness of estimations to the misspecification of \(N\) is an interesting point since the population size can be inaccurately known in practice. Assume that the true size of the population is \(N\) and that the wrong value \(N'\) has been used instead. Starting from the numbers of susceptible and infected individuals, we normalize by \(1/N'\) these quantities and then build the estimators. For the SIR model, we still obtain the right estimator for \(d\), while \(R_0\) is no longer consistently estimated. Our procedure estimates instead the quantity \(R_0N/N'\). This would also occur with estimation based on the ODE, while estimators based on all the jumps of \((S(t), I(t))\) would not exhibit this bias.

### 4.2 The SIRS model

For the SIRS model describing recurrent outbreaks and defined in Sect. 2.4, four parameters were estimated: in addition to \(R_0\) and \(d\), \(\lambda_1\) and \(\delta\) were assessed (the latter ones
Approximation of epidemic models by diffusion processes and their statistical inference

Fig. 4  Several simulated trajectories of the proportion of infected individuals over time using the SIR Markov jump process (top panels). Average point estimators (plus symbol) over 1,000 independent simulated trajectories (same model) and their associated theoretical confidence ellipsoids (bottom panels): MLE with complete observations (red), CE for 1 obs/day, n = 40 (blue), CE for n = 10 (black) and CE for n = 5 (green) for (R₀, d) = (1.5, 3) and N = [400, 1,000, 10,000] (from left to right). True parameter values are located at the intersection of horizontal and vertical dotted lines.

were replaced in estimations, for numerical reasons, by 10 × λ₁ and 1/δT_per. Demographic parameter μ was fixed to 1/50 years⁻¹, a value usually considered in epidemic models, T_per was taken equal to 365 days and η = 10⁻⁶, which corresponds to 10 individuals in a population size of N = 10⁷. The large value of N considered allows a sufficient pool of susceptible and infected individuals at the end of each outbreak for the epidemic to restart in the next season. Our CEs were assessed on trajectories obtained for (R₀, d, λ₁, δ) = {(1.5, 3, 0.05, 2), (1.5, 3, 0.15, 2)} and T = 20 years. The two scenarios correspond to λ₁ respectively smaller and larger than the bifurcation point of the corresponding ODE system (see Keeling and Rohani 2011, Chapter 5.1, for more details on the bifurcation diagram of the SIRS deterministic model). For numerical scenarios considered, the bifurcation value for λ₁ is around 0.07. However the qualitative pattern of epidemic dynamics (from annual to multiannual epidemics) also depends on the remaining parameter values (in particular, η seems to have an
Fig. 5  One simulated trajectory of the $SIR$ Markov jump process representing the simulated data (black line with crosses). Associated deterministic trajectories obtained with: the true parameter values, $(R_0, d) = (1.5, 3)$ (dotted black line), the MLE with complete observations $(1.46, 2.96)$ (red) and the CE $(1.56, 2.91)$ (blue).

Fig. 6  Deterministic (black) and Markov jump process (blue) trajectories of the $SIRS$ model with demography and seasonality in transmission. Proportion of infected individuals over time for $N = 10^7$, $(s_0, i_0) = (0.7, 10^{-4})$, $\eta = 10^{-6}$, $\mu = \frac{1}{50}$, $(R_0, d, \frac{1}{\Delta_{per}}) = (1.5, 3, 2)$, and $a \lambda_1 = 0.05$ and $b \lambda_1 = 0.15$.

important impact). As depicted in Fig. 6, for $\lambda_1 = 0.05$ the proportion of infectives exhibits oscillations which are roughly annual and of constant amplitude, whereas for $\lambda_1 = 0.15$ dynamics are biennial. Numerically, the scenarios considered have the characteristics of influenza seasonal outbreaks. According to results in Fig. 7, illustrating different projections of the four-dimensional theoretical confidence ellipsoid, almost no correlation is noticed between estimators, except for $R_0$ and $\lambda_1$ after bifurcation. Moreover, the accuracy of estimation is relatively high, regardless to the parameter. Interestingly, disposing of 1 obs/day (which can be considered as a practical limit of data availability) leads to an accuracy almost identical to the one corresponding to a complete observation of the epidemic process (blue and red ellipsoids respectively in

Springer
Fig. 7 Point estimators (plus symbol) built as averages over 1,000 independent simulated trajectories and their associated theoretical confidence ellipsoids for the SIRS model with demography and seasonal forcing in transmission: $CE$ for 1 obs/day (blue) and for $n = 1$ obs/week (black) for $(R_0, d, \lambda_1, \delta) = (1.5, 3, 0.15, 2), T = 20$ years and $N = 10^7$. Asymptotic confidence ellipsoid ($n \to \infty$) is also represented (red). True parameter values are located at the intersection of horizontal and vertical dotted lines.

Fig. 7). Estimations based on 1 obs/week provide less but still reasonably accurate estimations.

The width of $CI_{th}$ are similar for the two scenarios tested except for the parameter $\lambda_1$ (Fig. 7 for $\lambda_1 = 0.15$ and Figure S1 for $\lambda_1 = 0.05$). However, correlation between parameters (mainly $(R_0, \lambda_1)$ and $(d, 1/\delta T_{per})$) may vary according to the value of $\lambda_1$. The shape of the ellipsoid for $(d, 1/\delta T_{per})$ is also slightly different between the values tested and an SIRS model without seasonality (see electronic supplementary material, Figs. 7, Figure S1 and Figure S2). This can be partly explained by qualitative differences in corresponding deterministic dynamics of infected individuals. Finally, here again, one observation per day leads to remarkably accurate estimations.

5 Conclusion

In this study we provided first a rigorous and easy to implement three-step method for elaborating the diffusion approximation of Markov jump processes describing epidemic spread. Then, we developed a contrast-based inference method for parameters of epidemic models represented by diffusion processes, when all the coordinates of the system are discretely observed. The elaboration of the diffusion approximation built on results of Ethier and Kurtz (2005), but using a different technique as their time changed Poisson representation. Namely, we showed that the jump process and the diffusion both converge towards the same Gaussian process. Since our generalization encompasses time inhomogeneous systems, it allows handling complex epidemic models, particularly those with time-dependent transition rates. Our estimators have good properties for asymptotics corresponding to realistic situations in epidemiology, especially for large populations and limited number of observations. We also showed that a correction term introduced in the contrast function theoretically improved the
First, the process satisfying (4) is obtained recursively as follows. Let $Z_0(t) \equiv Z(0)$ and set $Z_1(t) = Z(0) + \sum_{l \in \mathbb{Z}_p} l \ P_l \left( \int_0^t \alpha_l(Z(0)) du \right)$. For $k > 1$, define $Z_k(t) = Z(0) + \sum_{l \in \mathbb{Z}_p} l \ P_l \left( \int_0^t \alpha_l(Z_{k-1}(u)) du \right)$. Then, if $\tau_k$ is the $k$th jump of $Z_k(t)$, $Z_k(t) = Z_{k-1}(t)$ for $t < \tau_k$. The process $Z(t) = \lim_{k \to \infty} Z_k(t)$ exists and satisfies (4).

Second, a characterization of these random time changed processes is mainly based on the property: given a positive measurable function $\eta : E \to (0, +\infty)$ and a Markov process $Y(.)$ such that $\int_0^\infty \frac{du}{\eta(Y(u))} = \infty$ a.s., one can define the random time change $\tau(t)$ by $\int_0^{\tau(t)} \frac{du}{\eta(Y(u))} = t \iff \tau(t) = \eta(Y(\tau(t)))$. The process $R(t)$ defined as $R(t) := Y(\tau(t))$ satisfies the equation $R(t) = Y(\int_0^t \eta(R(u)) du)$. Moreover, if $A$ is the generator of $Y(.)$, the generator of $R(t)$ is equal to $\eta A$. Now, if $(Y(t))$ is the Poisson process $P_l(t)$ with rate 1 (generator $Af(k) = f(k+1) - f(k)$), and $\eta(.) = \alpha_l(.)$, the process $Z_l(t) = P_l(\tau_l(t))$ has generator $A_l f(k) = \alpha_l(k)(f(k+1) - f(k))$ and satisfies, $Z_l(t) = P_l(\int_0^t \alpha_l(Z_l(s)) ds)$. This allows to prove that the solution of (4) has the generator $Af(k) = \sum_{l \in \mathbb{Z}_p} \alpha_l(k)(f(k+1) - f(k)) = \alpha(k) \sum_{l \in \mathbb{Z}_p} (f(k+1) - f(k)) \frac{\alpha_l(k)}{\alpha(k)}$. We identify this generator as the one of $(Z(t))$ defined by (2).

Acknowledgments Partial financial support for this research was provided by Ile de France Regional Council under MIDEM project in the framework DIM Malinf, and by French Research Agency, program Investments for the future, project ANR-10-BINF-07 (MIHMES).
A.2 Diffusion approximation for non-homogeneous Markov jump processes

We extend the approximation results from Ethier and Kurtz (2005) to the time dependent case. Their approach consists in using a Poisson time changed representation of the Markov jump process, a Brownian motion time changed representation of the diffusion process, and to compare them with an appropriate theorem from Komlós et al (1976). The extension of the proof of Ethier and Kurtz (2005) detailed in Appendix A.1 relies on the existence of (4) for time dependent Markov processes. The main problem is that the natural characterization of the random time change stated in Appendix A.1 now writes \( \int_0^t \frac{du}{n(\tau^{-1}(u), Y(u))} = t \), and the time change becomes implicit. We rather use the general convergence results from Jacod and Shiryaev (1987) to obtain the diffusion approximation.

We consider the pure Jump Markov process \( Z(t) \) with state space \( E = \{0, \ldots, N\}^p \) and transitions rates \( q_{x,x+i}(t) = \alpha_i(t,x) \). This process has for generator \( \mathcal{A}_t f(x) = \int_{\mathbb{R}^p} K_t(x, dy) (f(x+y) - f(x)) \) with the transition kernel \( K_t(x, dy) = \sum_{l \in E^-} \alpha_l(t,x) \delta_l(y) \) where \( \delta_l \) is the Dirac measure at point \( l \). Within the framework developed by Jacod and Shiryaev (1987), it is a semimartingale with a random jump measure integrating \( \| y \| \). So its characteristics in the sense of Definition 2.6 in Chapter II are \( (B, C, \nu_l) \), where

1. \( B = (B_i(t))_{1 \leq i \leq p} \) is the predictable process, \( B(t) = \int_0^t b(s, Z(s)) ds \), with \( b(s,x) = \int_{\mathbb{R}^p} y K_s(x, dy) = \sum_l \alpha_l(t,x) \).
2. \( C = (C_i(t)) \) is the quadratic variation of the continuous martingale part of \( Z(t) \), \( C(t) = (C_{i,j}(t))_{1 \leq i,j \leq p} \). For a pure jump process, \( C(t) \equiv 0 \).
3. \( \nu_l \) is the compensator of the jumps random measure of \( (Z_t) \), \( \nu_l(dt, dy) = dt K_t(Z(t), dy) = dt \sum_l \alpha_l(t, Z(t)) \delta_l(y) \).
4. The quadratic variation of the \( p \)-dimensional martingale \( M(t) = Z(t) - B(t) \) is, for \( 1 \leq i, j \leq p \), \( [M_{ij}] = \int_0^t m_{ij}(s) ds \) with \( m_{ij}(s) = \int_{\mathbb{R}^p} y_i y_j K_s(Z(s), dy) = \sum_l l_i l_j \alpha_l(t, Z(s))ds \).

Consider now the sequence of normalized pure jump processes \( Z_N(t) = \frac{Z(t)}{N} \) indexed by \( N \). The state space of \( Z_N \) is \( E_N = \{0, \frac{1}{N}, \ldots, 1\}^p \), its transition kernels are \( K_t^N(x, dy) = \sum_{l \in E^-} \alpha_l(t,Nx) \delta_{\frac{l}{N}}(dy) \). Hence, its characteristics are \( (B^N, C^N, \nu^N_l) \) with

1. \( B^N(t) = \int_0^t b^N(s, Z_N(s)) ds \), with \( b^N(s, x) = \sum_l \alpha_l(s, Nx) \frac{l}{N} \)
2. \( C^N(t) \equiv 0 \),
3. \( \nu^N_l(dt, dy) = dt K_t^N(Z_N(t), dy) = dt \sum_l \alpha_l(t,NZ_N(t)) \delta_{\frac{l}{N}}(y) \),
4. \( [M^N_{ij}] = \int_0^t m^N_{ij}(s) ds \), with \( m^N_{ij}(s) = \int_{\mathbb{R}^p} y_i y_j K^N_s(Z(s), dy) = \sum_l l_i l_j \alpha_l(s, NZ_N(s)) \frac{l}{N} \).

Under (H1), (H2), recall that \( b(t, x) = \sum_{l \in E^-} l \beta_l(t, x) \) and \( x_{00}(t) = x_0 + \int_0^t b(s, x_{00}(s)) ds \).

We first prove the convergence of the process \( Z_N(t) \) to \( x_{00}(t) \) (which has characteristics \((\int_0^t b(s, x_{00}(s)) ds, 0, 0)) \) by applying Theorem 3.27 of Chapter IX in Jacod and Shiryaev (1987). We have to check the following conditions:
(i) \( \forall t \in [0, T], \sup_{0 \leq s \leq t} \| B^N(t) - \int_0^t b(s, x_{x_0}(s)) ds \|_{N \to \infty} \to 0, \)

(ii) \( \forall t \in [0, T], [M^N_j](t) \to 0 \) in probability,

(iii) for all \( \eta > 0, \lim_{a \to +\infty} \limsup_{N} P \left\{ \int_0^1 \int_{\mathbb{R}^P} \| y \|^2 1_{\|y\| > a(y)} K_s^N(Z_N(s), dy) > \eta \right\} = 0, \)

(iv) \( \forall t \in [0, T], \int_0^t ds \int_{\mathbb{R}^P} y K_s^N(Z_N(s), dy) \to 0 \) in probability.

(v) \( Z_N(0) \to x_0 \) a.s.

Using (H1t), we obtain the uniform convergence of \( b^N(t, x) \to b(t, x) \) and \([M^N_j](t) \to 0 \) on \([0, T] \times [0, 1]^p\), which ensures conditions (i) and (ii). Condition (v) is satisfied by assumption. Since \( \int_{\mathbb{R}^P} \| y \|^2 K_s^N(x, dy) < \infty \), (iii) is satisfied. Using now that \( \int_{\mathbb{R}^P} \| y \| K_s^N(x, dy) < \infty \) yields (iv).

Therefore, \( Z_N(t) \to x_{x_0}(t) \) in distribution. Noting that \( b^N(t, x) \) and \([M^N_j](t) \) converge uniformly towards \( b(t, x) \) and 0 respectively, and using that the Skorokhod convergence coincides with the uniform convergence when the limit is continuous, we get

\[
sup_{t \in [0, T]} \| Z_N(s) - x_{x_0}(s) \|_{N \to \infty} \to 0 \text{ in probability.} \tag{21}
\]

It remains to study the process \( Y_N(t) = \sqrt{N} (Z_N(t) - x_{x_0}(t)) \).

For sake of clarity, we omit in the sequel the index \( x_0 \) in \( x_{x_0}(t) \). The jumps of \( Y_N \) have size \( 1/\sqrt{N} \), the transition kernel of the jumps random measure is \( \tilde{K}_t(y, du) = \sum_l \alpha_l(t, Nx(t) + \sqrt{N} y) \delta_{\frac{l}{\sqrt{N}}}(u) \), \( Y_N \) is a semimartingale with characteristics \((\tilde{B}^N, \tilde{C}^N, \tilde{v}^N)\)

1. \( \tilde{B}^N(t) = \int_0^t \tilde{B}^N(s, Y_N(s)) ds, \) with \( \tilde{B}^N(s, y) = \int_{\mathbb{R}^P} u \tilde{K}_s^N(y, du) - \sqrt{N} b(s, x(s)) = \sum_l \alpha_l(s, Nx(s) + \sqrt{N} y) \frac{l}{\sqrt{N}} - \sqrt{N} b(s, x(s)) \).

2. \( \tilde{C}^N(t) = 0. \)

3. \( \tilde{v}^N_t(dt, du) = dt \times \tilde{K}_t^N(Y_N(t), du) = dt \times (\sum_l \alpha_l(s, Nx(t) + \sqrt{N} Y_N(t)) \delta_{\frac{l}{\sqrt{N}}}(u)). \)

4. \( [\tilde{M}^N_{ij}](t) = \int_0^t \tilde{m}^N_{ij}(s) ds, \) with \( \tilde{m}^N_{ij}(s) = \int_{\mathbb{R}^P} u_i \times u_j \tilde{K}_s^N(Y_N(s), du) = \sum_l \alpha_l(s, N x(s) + \sqrt{N} Y_N(s)) \frac{l_i}{\sqrt{N}} \frac{l_j}{\sqrt{N}}. \)

Let us first study \( \tilde{B}^N(t) \). Using (H1) \( \frac{1}{\sqrt{N}} \alpha_l(t, N(x(t) + \sqrt{N})) = \beta_l(t, x(t) + \sqrt{N}) + r_N(t), \) with \( r_N(t) \to 0, \) by (H2), \( \beta_l(t, .) \) is differentiable and expanding \( \beta_l(t, .) \) around \( x(t) \) yields

\[
\beta_l(t, x(t) + \frac{\sqrt{N}}{l_{\sqrt{N}}}) = \beta_l(t, x(t)) + \sum_{i=1}^p \frac{\sqrt{N}}{l_i} \frac{\partial \beta_l}{\partial x_i}(t, x(t)) + \frac{1}{l_{\sqrt{N}}} r_{l_{\sqrt{N}}}(t), \text{ with } r_{l_{\sqrt{N}}}(t) \to 0.
\]

Therefore \( \frac{1}{\sqrt{N}} \alpha_l(t, N(x(t) + \sqrt{N})) = \sqrt{N} \beta_l(t, x(t)) + \sum_{i=1}^p \frac{\sqrt{N}}{l_i} \frac{\partial \beta_l}{\partial x_i}(t, x(t)) + \sqrt{N} r_N(t) + r_{l_{\sqrt{N}}}(t). \)

Hence, we need the additional assumption:
\begin{align}
(\text{H}1') \sqrt{N} \left( \frac{1}{N} \sum_{i=1}^{N} \alpha_i(t, Nx) - \beta_i(t, x) \right) & \to 0 \text{ uniformly w.r.t. } (t, x) \in [0, T] \times [0, 1]^p \text{ as } N \to \infty. \text{ Then, } \hat{B}^N(t, y) \to \sum_{i=1}^{P} y_i \int_0^1 \frac{\partial b_i}{\partial x_i}(t, x(s)) Y_N(s) ds.
\end{align}

Therefore, \( [\hat{M}^N_{ij}] (t) \to \int_0^t \sum_{i,j} \sigma(s, x(s)) ds \). Checking conditions (iii),(iv),(v) is straightforward. Finally, we obtain that \( Y_N \) converges in distribution to the process \( Y(t) \) with continuous sample paths, predictable process \( \int_0^t \nabla b(s, x(s)) Y(s) ds \) and quadratic variation \( \int_0^t \sum_{i,j} \sigma(s, x(s)) ds \). This is the diffusion process satisfying the SDE,

\begin{align}
dY(t) &= \left( \frac{\partial b_i}{\partial x_j} \right)_{i,j}(t, x(t)) Y(t) dt + \sigma(t, x(t)) dB(t); \quad Y(0) = 0,
\end{align}

where \( \sigma() \) satisfies \( \sigma(t, x)^\sigma(t, x) = \Sigma(t, x) \) and \( B(t) \) is a \( p \)-dimensional Brownian motion. This is an Ornstein-Uhlenbeck type SDE, which can be solved explicitly, leading to the Gaussian process \( G(t) \) previously introduced.

A.3 Extending the contrast approach (for non autonomous diffusion processes and for non constant sampling intervals)

Here, we provide the main line for the extension of the results in \textbf{Guy et al (2014)} for non autonomous diffusions and non constant sampling intervals. The complete proof is omitted. The main point of the proof of Proposition 3.2 in \textbf{Guy et al (2014)} relies on the relations (3.7) and (3.8) \( \frac{1}{\epsilon \Delta} A_k(\theta_0) \to V_k^{\theta_0} \) and \( \frac{1}{\Delta} \frac{\partial A_k(\theta)}{\partial \theta_j} \to D_{k,ij}(\theta_0) \). The proof of these relations is based on Taylor stochastic expansion and the fundamental relation of our contrast approach (18). The Taylor stochastic expansion of the diffusion was considered in \textbf{Freidlin and Wentzell (1978)} only for autonomous models, but has been extended for time dependent processes by \textbf{Azencott (1982)} and consequently holds when \( b \) and \( \Sigma \) are time dependent. Relation (18) is supported in the autonomous case by the semi-group property of function \( \Phi_{\theta} \) which leads to an associated analytic expression of \( g_{\theta}(t) = \int_0^t \Phi_{\theta}(t, s) \sigma(\theta, s) ds B_s \). Since the semi-group property is stated for non-autonomous cases in \textbf{Cartan (1971)}, the extension is immediate. For non constant sampling interval, the simple fact that relation (18) holds for any sequence \( t_0 < t_1 < \cdots < t_n \) ensures that the results of Proposition 3.2 in \textbf{Guy et al (2014)} hold.

References

\textbf{Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis. In: Lecture notes in statistics series. Springer, Berlin}

\textbf{Azencott R (1982) Formule de Taylor stochastique et développement asymptotique d’intégrales de Feynmann. Séminaire de Probabilités XVI, pp 237–285}

\textbf{Breto C, He D, Ionides EL, King AA (2009) Time series analysis via mechanistic models. Ann Appl Stat 3(1):319–348}

\textbf{Cao Y, Gillespie DT, Petzold LR (2005) Avoiding negative populations in explicit Poisson tau-leaping. J Chem Phys 123:054–104}

\textbf{Cartan H (1971) Differential calculus, vol 1. Hermann, Houston}

\textbf{Daley DJ, Gani J (2001) Epidemic modelling: an introduction. Cambridge University Press, Cambridge}
Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases: model building. Analysis and interpretation. Wiley, New York
Diekmann O, Heesterbeek H, Britton T (2012) Mathematical tools for understanding infectious disease dynamics. Princeton University Press, Princeton
Elerian O, Chib S, Shephard N (2001) Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69(4):959–993
Ether SN, Kurtz TG (2005) Markov processes: characterization and convergence, 2nd edn. Wiley, Fearnhead P, Papaspiliopoulos O, Roberts GO (2008) Particle filters for partially observed diffusions. J R Stat Soc Ser B 70(4):755–777
Freidlin M, Wentzell A (1978) Random perturbations of dynamical systems. Springer, Berlin
Fuchs C (2013) Inference for diffusion processes. Springer, Berlin
Genon-Catalot V (1990) Maximum contrast estimation for diffusion processes from discrete observations. Statistics 21(1):99–116
Genon-Catalot V, Jacod J (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. Annales de l’IHP Probabilités et statistiques 29(1):119–151
Genon-Catalot V, Larédo C (2006) Leroux’s method for general Hidden Markov models. Stoch Process Appl 116(2):222–243
Genon-Catalot V, Jeantheau T, Larédo C (2000) Stochastic volatility models as hidden markov models and statistical applications. Bernoulli 6(6):1051. doi:10.2307/3318471
Genon-Catalot V, Larédo C, Nussbaum M (2002) Asymptotic equivalence of estimating a Poisson intensity and a positive Diffusion Drift. Ann Stat 30(3):731–753
Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361
Gloter A (2001) Parameter estimation for a discrete sampling of an integrated Ornstein–Uhlenbeck process. Statistics 35(3):225–243
Gloter A, Sørensen M (2009) Estimation for stochastic differential equations with a small diffusion coefficient. Stoch Process Appl 119(3):679–699
Guy R, Larédo C, Vergu E (2014) Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. Stoch Process Appl 124(1):51–80
Jacod J, Shiryaev AN (1987) Limit theorems for stochastic processes. Springer, Berlin
Keeling MJ, Rohani P (2011) Modeling infectious diseases in humans and animals. Princeton University Press, Princeton
Kessler M (2000) Simple and explicit estimating functions for a discretely observed diffusion process. Scand J Stat 27(1):65–82
Komlós J, Major P, Tusnády G (1976) An approximation of partial sums of independent RV’s, and the sample DF II. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 34(1):33–58
Kutoyants YA (1984) Parameter estimation for stochastic processes. Heldermann, Lemgo
Larédo C (1990) A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. Ann Stat 18(3):1158–1171
Liptser RS, Shiryaev AN (2001) Statistic of random processes. Springer, Berlin
McKinley T, Cook AR, Deardon R (2009) Inference in epidemic models without likelihoods. Int J Biostat 5(1)
O’Neill PD (2010) Introduction and snapshot review: relating infectious disease transmission models to data. Stat Med 29(20):2069–2077
Ross JV, Pagendam DE, Polett PK (2009) On parameter estimation in population models II: multidimensional processes and transient dynamics. Theor Popul Biol 75(2–3):123–132
Sellke T (1983) On the asymptotic distribution of the size of a stochastic epidemic. J Appl Prob 20(2):390–394
Sørensen M, Uchida M (2003) Small-diffusion asymptotics for discretely sampled stochastic differential equations. Bernoulli 9(6):1051–1069
van der Vaart A (2000) Asymptotic statistics. Cambridge University Press, Cambridge
Van Kampen NG (1992) Stochastic processes in physics and chemistry. Elsevier, Amsterdam