ROMA TRE
UNIVERSITÀ DEGLI STUDI

SCUOLA DOTTORALE IN ECONOMIA E METODI QUANTITATIVI - METODI STATISTICI PER L'ECONOMIA E L'IMPRESA

XXIV CICLO SECS-S/01

Stochastic Bass diffusion: model properties, exact simulation and statistical inference

Dottorando: **Ugo Guarnera**

Tutor: **Prof. Giovanni Sebastiani**

Coordinatore: **Prof.ssa Julia Mortera**

A Fillimar e Marzietta

Ringrazio il Prof. Giovanni Sebastiani per i suggerimenti e i consigli elargiti durante tutto il periodo della stesura della tesi.

Ringrazio la Prof.ssa Julia Mortera per l'organizzazione e il cordinamento delle attività del Dottorato e per il vigoroso incoraggiamento.

Ringrazio Teresa Buglielli, Marco Di Zio e Roberta Varriale per la preziosa assistenza tecnica che mi ha consentito di superare in tempi ragionevoli problemi piccoli e meno piccoli.

Ringrazio i miei colleghi di ufficio per avermi tollerato quando, sotto la duplice pressione degli impegni di lavoro e di studio, ho inflitto loro tediose lamentele e reazioni abnormi a normali avversità. Li ringrazio anche per aver subito le conseguenze del mio sensibile calo di produttività.

Ringrazio Arianna per aver sopportato il peso della convivenza con un attempato studente lavoratore, e per avermi fornito un supporto materiale e psicologico indispensabile per portare a compimento l'impresa.

Rinrazio me stesso per essere stato in grado di trovare motivazioni e stimoli nei momenti più difficili resistendo alla tentazione di mandare tutto in malora.

Ugo

Contents

List of Figures

List of Tables

Introduction

The need for modelling diffusion of innovations in a statistical framework has been motivating important reasearch from both practical and theoretical perespective for the last 30 years. Diffusion of innovation has been analyzed both at a microscopic level (Niu, 2002; R.Guseo and M.Guidolin, 2010; Moldovan and J.Goldenberg, 2003) and at an aggregate level (Fourt and Woodlock, 1960; Bass, 1969; Mansfield, 1961; Rogers, 1995; Mahajan and Wind, 1986; Skiadas and Giovanis, 1997). In the microscopic approach, what is modeled is the behaviour of a single individual, interacting with other individuals of a social system, in terms of adoption of a recently introduced innovation. At a given time, the probability of adopting the innovation is associated to each individual. Depending on the model complexity, the adoption probability depends on a certain number of factors, such as the specific characteristics of the potential adopter, or the environmental conditions. Typically, the probability for an individual to adopt the innovation depends on the number of people who are already adopters: the higher the number of adopters, the higher the probability of a new adoption. In this respect, the diffusion of innovations is similar to other diffusion phenomena, such as the diffusion of a desease, or the dynamics of a population, and can be studied through the same methodologies, (e.g., birth-death processes, SIR/SIS models).

With the aggregate approach, the dynamics of the total number of adopters, instead of individual behaviour, is directly modeled. In the approaches based on the classical Bass model (Bass, 1969), cumulative adoption is modeled through a deterministic differential equation of logistic-type, where the instant rate of adopter number at time t is assumed to be proportional to the sum of two terms. The first term is the adopter number and accounts for the individual propensity to the adoption ("self-innovation"). The second term is the product of the number of people who at time t are already adopters and the ones who are not. This term, considered the most important in many applications, captures the "word by mouth" or "imitation" effect, that takes

into account, under suitably simplifying assumptions, the impact of the communication among the individuals on the adoption. The relationship between micro and aggregate approach is, in general, an interesting field of research. In particular, it is of great interest to analyze under which conditions models for cumulative adoption can be derived from probabilistic micro-level models. For instance, in the limit of large size population, it can be proved (Niu, 2002) that the deterministc Bass equation is fulfilled by the expected value of the total adopter number, under a micro-level dynamic governed by a pure birth-death process.

In this thesis, the aggregate approach is adopted. In particular, the classical Bass model is generalized by introducing random noise in the dynamics of the diffusion. Stochastic aggregate models for the diffusion of innovation have been studied by many authors. For instance, Gutierrez et al. (2005) use an extension of the classical Gompertz innovation diffusion model to predict natural-gas consumption in Spain. Kanniainen et al. (2011) and Skiadas and Giovanis (1997) propose two stochastic extensions of the Bass model: in the first paper a mean reverting Orstein-Uhlenbeck process is added to the logarithm of the classical solution of the Bass equation, while in the second the authors propose an SDE which is obtained from the classical Bass equation by adding a Brownian diffusion term chosen so that an explicit analytic solution of the SDE can be found. Introducing randomness in the diffusion Bass model by replacing the ordinary differential equation (ODE) with a suitable stochastic differential equation (SDE) is a natural approach. The crucial point is the choice of the diffusion coefficient. In fact, in order to interpret the solution of the SDE as a (continuous approximation of) stochastic process corresponding to the number of adopters, some qualitative properties have to be satisfied. For instance, it is desirable that at each time the process takes a.s. values belonging to the interval $[0, K]$, where K is the number of potential adopters (*regularity conditions*). This latter condition is not satisfied, for instance, if the diffusion coefficient is chosen as in Skiadas and Giovanis (1997).

In this thesis, the diffusion term has been defined so that the resulting stochastic Bass model has the right regularity conditions. Moreover, the drift of the SDE includes a term which accounts for the possibility that some people cease being adopters ("disadoption"). The resulting stochastic Bass model (SBM) is a parametric model with four "deterministic" parameters associated, respectively, to: self-innovation (a), imitation (b), disadoption (μ) and number of potential adopters (K), and one "stochastic" parameter (σ) governing the volatility of the process. As disadoption is allowed, and in

any case, since random noise can determine decrease of adoptions, the SBM is not an a.s. increasing process. This makes it suitable to model adoption of innovations referring to services or social behaviours, more than diffusion of new products measured in terms of sale growth. However, the SBM can be useful in modeling sales period by period (for instance daily sales), for kind of durables whose consumption is believed to be approximately "proportional" to the number of users.
Using some results on the regularity of solution of logistic-type SDEs in Shurz (2007), it can be proved that sample paths from the law of the SBM remain in the interval $[0, K]$ provided that the process starts in the same interval and that all model parameters are non negative. In this thesis (Chapter 1), the results on regularity conditions have been extended to include also the (time non-homogeneous) case where the number of potential adopters is a deterministic non decreasing function of time. This extension can be useful when the effects of the population dynamics on the diffusion of innovation are to be taken into account.

In this thesis, some theoretical properties of the SBM have been explored. It has been proved that when all the model parameters are positive, an equilibrium distribution exists (i.e., the process is ergodic). Furthermore, its explicit representation in terms of modified Bessel functions of second kind has been derived. In case of no self-innovation ($a = 0$), the conditions on the other parameters for the existence of the equilibrium distribution have been investigated. Moreover, the stochastic stability of the trivial solutions corresponding to "no adopter" or "all adopters" has been studied, when the parameters are such that the process has not a proper stationary distribution.

Although the stochastic Bass model has good theoretical properties that make it suitable to model diffusion phenomena, the analytic solution of the underlying SDE, as well as the explicit form of its transition density, is not available. This poses serious problems in statistical inference for the SBM parameters, based on a finite set of observations.

Inference on discretely observed diffusion processes is a challenging area of research. In fact, except for a few cases (e.g., solutions of linear SDE or CIR model) the transition density of the diffusion process cannot be expressed in analytic form. Hence, likelihood-based inferential approaches, such as maximum likelihood (ML) estimation and Bayesian inference, are not easy to apply because the likelihood function is not known. Unavailability of the transition density makes also difficult to evaluate specific functionals of the diffusion process, such as hitting times, or expectations at given times. In

these cases, one possible approach is to simulate a large number of process realizations and to approximate the quantities of interest via sample averages. A common method for random drawing of approximate realizations from the law of a diffusion process is to discretize the time interval on which the process is defined and to approximate the original SDE via some Eulerian-type scheme. As the discretization step tends to zero, the solution of the approximate SDE converges, in a suitable metrics, to the solution of the original SDE. Thus, the accuracy of the approximations depends on a further parameter (the discretization step), in addition to the number of MC replicates used to approximate expectations. Another problem with Eulerian methods is that boundness conditions, such as regularity properties with respect to a given domain, are not ensured. For instance, in the case of the SBM model, there is no guarantee that the process values remain lower than K, the number of potential adopters.

In this thesis, it is shown how sample paths from the "exact" law of the stochastic Bass model without self-innovation (SBM1) can be drawn. Exact simulation of SBM1 trajectories is based on the methodology of *retrospective sampling* recently introduced by Beskos, Papaspilioupolos and Roberts (2006). Essentially, it is an Acceptance-Rejections (A/R) scheme which allows, for a cetain class of diffusions, to accept (or reject) a proposal in an infinite-dimensional space, by checking only a finite set of conditions. Besides "exactness", another important advantage of the exact simulation methodology for diffusion processes, is that, differently from the Eulerian methods, it also allows one to easily obtain realizations from the law of the process conditioned on both its starting and ending points (*diffusion bridge*). This characteristic is of great importance in all those inferential methods, as EM algorithm or Gibbs sampling, based on data augmentation schemes. In this thesis, however, major emphasis is posed on inferential methods based on Monte Carlo (MC) approximation of the transition density. Specifically, for two given consecutive observations of the SBM process, one starts by considering the diffusion bridge obtained by conditioning the target process to these two observations and the times they refer to. Then, exploiting the expression of the Radom Nikodyn derivative of the target diffusion bridge with respect to the corresponding Brownian bridge, one can express the transition density in terms of expectation of a suitable Brownian functional. This expectation is analytically intractable but can be easily approximated via MC average. MC approximation of the transition density for each pair of consecutive obervations provides, in turn, an MC approximation of the likelihod function. Thus, approximate maximum likelihood estimates (*amle*) of the model parameters can be obtained by maximizing the approximate likelihood. The *amle* methodology has been applied to the estimation of the

SBM1 model parameters and compared with other classical methodologies. The comparison shows that *amle* performs better than the competing techniques, especially when the process is sampled at low frequency.

The thesis is organized as follows. In Chapter 1 the stochastic Bass model is introduced. The new results on the regularity conditions and the derivation of the stationary distribution are presented in Subsections 1.2.2 and 1.2.3 respectively. The latter section also contains the analysis of stochastic stability of the degenerate equilibrium solutions of the Bass SDE. In Chapter 2 the methdology of exact simulation of diffusion processes based on retrospective sampling is illustrated. Chapter 3 contains a description of different approaches to the parametric inference for discretely observed diffusions. In particular, a detailed description of the inferential methods based on MC approximation of the transition density is provided. Chapter 4 is devoted to exact simulation and statistical inference for the stochastic Bass model. The methodologies introduced in general in Chapters 2 and 3 are described for the case of SBM in Sections 4.1 and 4.2 respectively. In particular, in Section 4.2 (Subsection 4.2.3), a numerical application is presented where the *amle* is compared with the classical estimator based on the Gaussian approximation. Finally, comments and remarks are presented in a concluding section.

Chapter 1

The stochastic Bass diffusion

1.1 The classical Bass model

In marketing sciences, a growing area of research is focused on market demand for new products or services. In this field, it is of great interest to model the diffusion of *innovation* within a population of potential consumers/users in order to optimize promotion and distribution strategy. To this aim, many deterministic models have been proposed in the last three decades to capture the most important aspects of innovation diffusion. See, for instance, Bass (1969), Mahajan and Wind (1986), Parker (1994), Rogers (1995). The dynamics of the population of interest is often modeled by an ordinary differential equation (ODE). In most of these works, efforts are made in order to properly take into account the role played by the communication among the members of a network or social system. Effects on innovation diffusion of contacts among individuals of the population are typically captured by nonlinear terms in the ODE. Logistic-type equations, often used for modeling population growth, are particulary suitable to model diffusion of innovation. Among them, the Bass model (Bass, 1969) is often used:

$$\frac{dy}{dt} = a(K - y) + b\frac{y}{K}(K - y) - \mu y, \tag{1.1}$$

where $y(t)$ represents the number of innovation *adopters* at time t. The three terms on the r.h.s. of equation (1.1) can be interpreted as follws: The first one (*self-innovation*) models the effect of commercials on adoption and is proportional to the number $K - y(t)$ of individuals who are not adopters at time t, where the positive parameter K is the number of potential adopters. The second term is sometimes called "imitation" or "word of mouth" and captures the effect of interactions between individuals on adoption. The last

term takes into account the effect of adopters who stop being so. All the parameters a, b, μ are non negative. In some contexts, the second term can also be directly related to the benefit for new adopters due to the presence of current adopters. This is the case for instance, when adoptions refer to goods or services (e.g., social networks) such that individual utility increases with the number of current adopters.

If $(b-a-\mu)^2+4ab>0$ equation (1.1) has the two equilibrium points:

$$y^{\pm}=K\frac{b-a-\mu\pm\sqrt{(b-a-\mu)^2+4ab}}{2b}. \tag{1.2}$$

It is easy to verify that $y^+\in[0,K]$ and is asymptotically stable, while y^- is non positive and unstable. The non-negative solution of equation (1.1) with initial condition $y(0)=y_0\geq 0$ is:

$$y(t)=\frac{y^+A\exp\left\{\frac{b}{K}(y^+-y^-)t)\right\}+y^-}{1+A\exp\left\{\frac{b}{K}(y^+-y^-)t\right\}}, \tag{1.3}$$

where $A=(y_0-y^-)/(y^+-y_0)$. Furthermore, one can easily check that, for $t>0$, $y'(t)$ is positive if $y_0<y^+$, and negative if $y_0>y^+$. It follows that for $y_0<y^+$ ($y_0>y^+$), the solution (1.3) is an increasing (decreasing) positive function with $\lim_{t\to\infty}y(t)=y^+$. Thus, starting from the initial value y_0, the number of adopters tends to its asymptotic value y^+, i.e., the system goes to "saturation". It is worthwhile noting that, if mortality is not included in the model ($\mu=0$), as for instance in the case when y is the number of purchases of a certain good, then $y^+=K$, i.e., for t large enough, in practice all potential consumers become adopters. It is of some interest to measure the speed at which the system goes to saturation. A frequent parameter used to this aim is the time t^* at which $y'(t)$ is maximum (*inflaction time*). Direct calculation shows that

$$t^*=\frac{K}{2b(y^+-y^-)}\ln\frac{y^+-y_0}{y_0-y^-}.$$

In the case $(b-a-\mu)^2+4ab=0$, i.e., $a=0$, $b=\mu$, the two stationary points collapse in the point $y^*=0$ and the solution of (1.1) is:

$$y(t)=\frac{1}{y_0^{-1}+\frac{b}{K}t}.$$

1.2 The stochastic Bass model

Deterministic model (1.1) can be justified in different ways. For example, in Niu (2002) it is shown that the evolution law for the proportion $y(t)/K$, provided by equation (1.3), can be derived, in the limit of large population size, from a stochastic microscopic model where the trajectory of cumulative number of adoptions is governed by a pure birth process. In R.Guseo and M.Guidolin (2010), the authors obtain the solution (1.3) from a suitable Cellular Automata representation under the mean field approximation.
Recently, there has been a growing interest in the stochastic generalization of equation (1.1) capable of taking into account the presence of random mechanisms in the adoption dynamics. This allows to have non-monotone solutions as it is the case in practice in some applications. In the next Section, a stochastic generalization of the Bass model will be introduced and some of its known theoretical properties will be described.

1.2.1 Introduction and known properties

A natural way of introducing stochasticity in the model is to replace the deterministic equation (1.1) with an appropriate stochastic differential equation (SDE):

$$dY_t = \left[a(K - Y_t) + b\frac{Y_t}{K}(K - Y_t) - \mu Y_t \right] dt + \tilde{\sigma}(Y_t) dW_t, \tag{1.4}$$

where $\tilde{\sigma}(\cdot)$ is the volatility function and W_t is a one-dimensional Brownian motion. For instance, this is the approach of Skiadas and Giovanis (1997), who propose a particular choice for $\tilde{\sigma}$. While the choice of the drift term is naturally guided by the deterministic version of the differential equation, it is not obvious how to specify the stochastic term. As we will see soon, some restrictions are imposed by the requirement that the solutions have certain reasonable qualitative properties. In particular, since y represents the number of adopters, it must remain non-negative. Furthermore, we wish the number of adopters does not exceed the maximum number K of potential adopters. In other words, we want that the trajectory evolves within the interval $[0, K]$. According to the terminology of Shurz (2007), we call the process Y_t *regular* with respect to the domain $\mathcal{D} \subset \mathbb{R}$, if $P\{Y(t) \in \mathcal{D}; t > 0\} = 1$.

The following theorem (Shurz, 2007) shows that the solution of (1.4) is not regular on $[0, K]$ if $\tilde{\sigma}$ is a numerical constant.

Theorem 1.2.1 *Assume that* $\{Y_t, t \geq 0\}$ *satisfies equation (1.4) with* $\tilde{\sigma} = \sigma > 0, K > 0, a \geq 0, b \geq 0, \mu \geq 0$, *and* $Y_0 \in [0, K]$ *is independent of the* σ*-algebra* $\sigma(W_t, t \geq 0)$. *Then* $\{Y_t, t \geq 0\}$ *is not regular with respect to* $\mathcal{D} = [0, K]$. *More precisely,* $\forall x \in \mathcal{D}$, $P(\tau^x(\mathcal{D}) < +\infty) > 0$, *where* $\tau^x(\mathcal{D})$ *is the first exit time of the process* Y *from the domain* $\mathcal{D}$, *when the starting point is* x.

The previous result shows that although the drift has the "right sign" on the boundary of the region $\mathcal{D}$, in the classical sense that it pushes the motion towards the interior of the region, additive random noise may cause the process to leave the region in finite time. It is reasonable to expect that, in order to have regularity with respect to region $\mathcal{D}$, the diffusion term has to vanish on $\partial\mathcal{D}$. The next result (Shurz, 2007) shows that regularity is ensured in a class of SDEs with "multiplicative noise".

Theorem 1.2.2 *Let* $\{Y_t, t \geq 0\}$ *be a strong solution of the Bass SDE :*

$$dY_t = \left[a\left(K - Y_t\right) + b\frac{Y_t}{K}\left(K - Y_t\right) - \mu Y_t\right]dt + \frac{\sigma}{K}|Y_t|^{\alpha}|K - Y_t|^{\beta}dW_t, \quad (1.5)$$

with initial condition

$$Y_0 \in (0, K),$$

where $K \geq 1$, *and* $a, b, \mu, \sigma, \alpha, \beta$ *are non-negative constants.*

Let τ *be the stopping time defined by:*

$$\tau = \inf\left\{t > 0 : Y_t \notin (0, K)\right\}.$$

Then, under the conditions $\alpha \geq 1$, $\beta \geq 1$, *we have* $P(\tau < t) = 0$, $\forall t > 0$.

1.2.2 Extension and new properties

In this section, we prove a new result corresponding to a generalization of the last theorem. Specifically, we assume that K is no more constant, but it is varying along time, i.e., $K \equiv K(t)$. The extendend non homogeneous Bass equation is:

$$dY_t = B(Y_t, t)dt + D(Y_t, t)dW_t, \quad (1.6)$$

where $B(\cdot,\cdot)$ and $D(\cdot,\cdot)$ denote the drift and the diffusion coefficients, respectively, defined by:

$$B(y,t) \doteq a\,(K(t)-y) + b\frac{y}{K(t)}\,(K(t)-y) - \mu y; \tag{1.7}$$

$$D(y,t) \doteq \frac{\sigma}{K(t)}|y|^{\alpha}|K(t)-y|^{\beta}. \tag{1.8}$$

Theorem 1.2.3 *Let $\{Y_t, t \geq 0\}$ be a strong solution of equation (1.6) with drift and diffusion functions (1.7), (1.8), and initial condition:*

$$Y_0 \in (0, K(0)), \qquad K(0) > 1.$$

Assume that $K(\cdot) \in C^1(\mathbb{R}^+ \cup \{0\})$ is a non decreasing function and $a, b, \mu, \sigma, \alpha, \beta$ are as in theorem (1.2.2). Then, if τ is the stopping time defined by:

$$\tau = \inf\{u > 0 : Y_u \notin \mathcal{D}(u)\},$$

where $\mathcal{D}(t) \doteq (0, K(t))$.

we have $P(\tau < t) = 0,\ \forall t > 0$.

Proof. Let $T > 0$ and $K^* = \max\{K(t),\ t \in [0,T]\}$. For each $t \in [0,T]$ and $n \in \mathbb{N}$ define the "time dependent" open interval:

$$\mathcal{D}_n(t) = (e^{-n}, K(t) - e^{-n}).$$

Define the random time: $\tau_n = \inf\{u \in [0,T] : Y_u \notin \mathcal{D}_n(u)\}$. We assume that $\tau_n = +\infty$, if $\{u \in [0,T] : Y_u \notin \mathcal{D}_n(u)\} = \emptyset$. For $t \in [0,T]$, let us introduce the Lyapunov function V:

$$V(x,t) = K^* - \log(x(K(t)-x)).$$

From the elementary inequality $y - 1 \geq \log y$ valid for all positive y, it follows that:

$$V(x,t) = K^* - K(t) + K(t) - x - \log(K(t)-x) + x - \log x \geq K^* - K(t) + 2 \geq 2.$$

Let us introduce a second Lyapunov function W defined as:

$$W(x,t) = e^{-c(t)t} V(x,t),$$

where:

$$c(t) \doteq \frac{a + b + \sigma^2 K(t)^{2\alpha+2\beta-4} + \mu}{2}.$$

Now, for fixed $t \in [0,T]$, we introduce another stopping time $\tau_n^* = t \wedge \tau_n$ and we use the Dynkin formula (see, e.g., Oksendal (1998)) to obtain an expression of the expectation of W computed at the (random) point $(Y_{\tau_n^*}, \tau_n^*)$:

$$E(W(Y_{\tau_n^*}, \tau_n^*)) = W(Y_0, 0) + E\left\{ \int_0^{\tau_n^*} \left[\frac{\partial W}{\partial u}(Y_u, u) + \mathcal{L}_0 W(Y_u, u) \right] du \right\}, \quad (1.9)$$

where:

$$\mathcal{L}_0 \doteq B \frac{\partial}{\partial y} + \frac{1}{2} D^2 \frac{\partial^2}{\partial y^2}. \quad (1.10)$$

The integrand in (1.9) can be written as:

$$-cW - c'uW + \exp(-c(u)u) \frac{\partial V}{\partial u} + \exp(-c(u)u) \mathcal{L}_0 V, \quad (1.11)$$

where the simbol $'$ denotes the derivative with respect to time. Now, we prove the following inequality:

$$\mathcal{L}_0 V \leq cV. \quad (1.12)$$

Note that, for fixed $t \in [0,T]$ and $0 < y < K(t)$, from $V(y,t) \geq 2$ it follows that:

$$
\begin{aligned}
cV(y,t) - \mathcal{L}_0 V(y,t) &\geq a + b + \sigma^2 K^{2\alpha+2\beta-4} + \mu - \mathcal{L}_0 V(y,t) \\
&= a + b + \sigma^2 K^{2\alpha+2\beta-4} + \mu - \Big[(a + b\frac{y}{K})(K-y) - \mu y\Big] \\
&\quad \times \left[-\frac{1}{y} + \frac{1}{K-y}\right] - \frac{\sigma^2}{2K^2} y^{2\alpha}(K-y)^{2\beta}\left[\frac{1}{y^2} + \frac{1}{(K-y)^2}\right] \\
&\geq \sigma^2 K^{2\alpha+2\beta-4} + \frac{(a + by/K)(K-y)}{y} + \frac{\mu y}{K-y} \\
&\quad - \frac{\sigma^2}{2K^2} y^{2\alpha-2}(K-y)^{2\beta-2}\left[(K-y)^2 + y^2\right] \\
&\geq \sigma^2 K^{2\alpha+2\beta-4} - \frac{\sigma^2}{2K^2} y^{2\alpha-2}(K-y)^{2\beta-2}\left[(K-y)^2 + y^2\right] \\
&\geq \sigma^2 K^{2\alpha+2\beta-4}\left[1 - \frac{(K-y)^2 + y^2}{2K^2}\right] \geq 0,
\end{aligned}
$$

where, in order to make notation simpler, we have omitted the argument t in the functions c and K. From (1.9), (1.11) and (1.12), when $Y_u \in (0, K(u))$ for $u \in (0, \tau_n^*)$, and $\alpha, \beta \geq 1$, it follows:

$$
\begin{aligned}
E(W(Y_{\tau_n^*}, \tau_n^*)) &\leq V(Y_0, 0) + E\left\{\int_0^{\tau_n^*}\left[-c'uW + \exp(-c(u)u)\frac{\partial V}{\partial u}\right] du\right\} \\
&\leq V(Y_0, 0) + E\left\{\int_0^{\tau_n^*} \exp(-c(u)u)\frac{\partial V}{\partial u} du\right\} \\
&= V(Y_0, 0) + E\left\{\int_0^{\tau_n^*} \exp(-c(u)u)\left[-\frac{1}{K(u) - Y_u}\right] K'(u) du\right\} \\
&\leq V(Y_0, 0),
\end{aligned}
$$

where the first and last inequalities hold because $K(\cdot)$ is not decreasing.

Now, it is easily seen that for $t > 0$, it holds:

$V(\cdot, t) > 1 + n$ on $\mathcal{D}(t) \setminus \mathcal{D}_n(t)$. Thus, $V > 1 + n$ on the open set

$$U_n \doteq \left\{(y,s):\ K(s)-e^{-n}<y<K(s),\ 0\le s\le t\right\}\cup\left\{(y,s):\ 0<y<e^{-n},\ 0\le s\le t\right\}.$$

For fixed $t<T$ we have:

$$P(\tau<t)\le P(\tau_n<t)=P(\tau_n^*<t)=E(\mathbf{1}_{\tau_n^*<t}). \tag{1.13}$$

Using the inequality $E(W(Y_{\tau_n^*},\tau_n^*))\le V(Y_0,0)$, previously proved, and the fact that $c(t)$ is positive and not decreasing, we obtain:

$$\begin{aligned}
P(\tau<t)\le E(\mathbf{1}_{\tau_n^*<t}) &\le E\left\{\frac{\exp\left[c(\tau_n^*)(t-\tau_n^*)\right]V(Y_{\tau_n^*},\tau_n^*)}{\inf\left\{V(x,s):(x,s)\in U_n\right\}}\mathbf{1}_{\tau_n^*<t}\right\}\\
&\le E\left\{\frac{\exp\left[c(\tau_n^*)(t-\tau_n^*)\right]V(Y_{\tau_n^*},\tau_n^*)}{\inf\left\{V(x,s):(x,s)\in U_n\right\}}\right\}\\
&\le e^{c(t)t}E\left\{\frac{\exp\left[-c(\tau_n^*)\tau_n^*\right]V(Y_{\tau_n^*},\tau_n^*)}{\inf\left\{V(x,s):(x,s)\in U_n\right\}}\right\}\\
&\le e^{c(t)t}\frac{V(Y_0,0)}{1+n}\overset{n\to\infty}{\longrightarrow}0.
\end{aligned}$$

Since times t and T are arbitrary, the thesis follows.

1.2.3 Stationary distribution and asymptotic stability

In this section, we cope with the particular case $\alpha=1,\beta=1$, and K constant. We refer to this model as SB model (shortly, SBM) and the corresponding stochastic process as SBM process. We notice that the result proven in the previous section allows us to remove the absolute values appearing in the definition of the stochastic term (1.8). Now we are interested in analyzing conditions for the parameters a,b,μ,σ that ensure the existence of equilibrium distribution. When the equilibrium distribution does not exist,

we study the asymptotic stability of the singular equilibrium solutions $Y_t \equiv 0$ and $Y_t \equiv K$. Furthermore, hereinafter we will assume that the parameter b is strictly positive. In fact, the condition $b = 0$ would imply the absence of the "word of mouth" effect, and the corresponding process is of no interest for modeling diffusion of innovations. Moreover, the corresponding SDE would become linear. Let Y_t be the solution of the SDE (1.5) with (deterministic) initial condition $Y_0 \in (0, K)$ and volatility function

$$\tilde{\sigma} = \frac{\sigma}{K} Y_t \left(K - Y_t\right),$$

i.e., of the SDE:

$$dY_t = \left[a(K - Y_t) + \frac{b}{K} Y_t(K - Y_t) - \mu Y_t\right] dt + \sigma \frac{Y_t}{K} \left(K - Y_t\right) dW_t. \quad (1.14)$$

First, we note that, the "proportion" process defined as:

$$Z_t = \frac{Y_t}{K}$$

is a strong solution of the SDE:

$$dZ_t = \left[a(1 - Z_t) + bZ_t(1 - Z_t) - \mu Z_t\right] dt + \sigma Z_t \left(1 - Z_t\right) dW_t, , \quad (1.15)$$

with initial condition:

$$Z_0 = \frac{Y_0}{K} \in (0, 1).$$

It is convenient to transform the process into a new process with unit volatility. To this aim, we use the Lampert transformation and we introduce the new process valued on the entire real axis:

$$X_t \doteq \eta(Z_t) \doteq -\frac{1}{\sigma} \log \frac{Z_t}{1 - Z_t}.$$

From the Ito Lemma, it follows that the process X is solution of the SDE:

$$dX_t = \alpha(X_t)dt + dW_t, \quad (1.16)$$

where the new drift $\alpha(\cdot)$ is defined as:

$$\begin{aligned} \alpha(x) &= D - \frac{a}{\sigma}e^{\sigma x} + \frac{\mu}{\sigma}e^{-\sigma x} - \sigma\frac{e^{-\sigma x}}{1+e^{-\sigma x}}, \\ D &= -\frac{a}{\sigma} - \frac{b}{\sigma} + \frac{\mu}{\sigma} + \frac{\sigma}{2}. \end{aligned}$$

To find the stationary distribution density $q(\cdot)$ of the process X_t, we look for a solution of the homogenous Fokker Plank equation (Risken, 1996):

$$\frac{1}{2}q''(x) - \frac{d}{dx}\left(\alpha(x)q(x)\right) = 0. \tag{1.17}$$

After integrating equation (1.17) we have:

$$\frac{1}{2}q'(x) - \left(\alpha(x)q(x)\right) = C, \tag{1.18}$$

where C is a constant. The general solution of equation (1.18) can be written as:

$$\begin{aligned} q(x) &= K_1 e^{2A(x)} + 2Ce^{2A(x)}\int_d^x e^{-2A(t)}dt \\ &= e^{2A(x)}\left[K_1 + 2C\int_d^x e^{-2A(t)}dt\right], \end{aligned} \tag{1.19}$$

where

$$A(x) = \int_d^x \alpha(t)dt = Dx - \frac{a}{\sigma^2}e^{\sigma x} - \frac{\mu}{\sigma^2}e^{-\sigma x} + \log(1+e^{-\sigma x}) - K_2,$$

K_1 and d are constants such that $0 < q(d) = K_1$, and $K_2 = Dd - \frac{a}{\sigma^2}e^{\sigma d} - \frac{\mu}{\sigma^2}e^{-\sigma d} + \log(1+e^{-\sigma d})$.

In order for $q(\cdot)$ to be a probability density, it must be positive everywhere on $(-\infty, +\infty)$ and tend to zero as $x \to \pm\infty$. As we will see, this implies some restrictions on the family of solutions (1.19). Let us consider different cases separately.

i) $\mathbf{a > 0},\ \mu > \mathbf{0}$ (both self-innovation and recess are possible).

we have:

$$\lim_{x\to\pm\infty} A(x) = \lim_{x\to\pm\infty} \left[Dx - \frac{a}{\sigma^2}e^{\sigma x} - \frac{\mu}{\sigma^2}e^{-\sigma x} + \log(1+e^{-\sigma x}) - K_2\right] = -\infty$$

The requirement of non negativity for $q(\cdot)$ implies that $C = 0$. In fact, for $C \neq 0$, the quantity $1 + (2C/K_1)\int_d^x e^{-2A(t)}dt$, for $|x|$ large enough, would have different sign for different sign of $x - d$, hence $q(\cdot)$ would fail to be positive on the entire real axis.

Thus, the stationary distribution density $q(\cdot)$ reduces to:

$$q(x) = K_1 e^{2A(x)} = \mathcal{Z}^{-1} \exp\left\{2Dx - 2\frac{a}{\sigma^2}e^{\sigma x} - 2\frac{\mu}{\sigma^2}e^{-\sigma x} + 2\log(1+e^{-\sigma x})\right\},$$

where $\mathcal{Z}$ is the normalization constant. In order to obtain the equilibrium distribution for the original process Y_t, we use the relation:

$$x = \eta(y/K) = -\frac{1}{\sigma}\log\frac{y}{K-y}.$$

Thus, the stationary distribution density $p(y)$ for the original process is:

$$\begin{aligned} p(y) &= q(\eta(y/K))\frac{1}{K}\left|\frac{d\eta}{dz}\right|_{z=y/K} \\ &= \mathcal{Z}^{-1}\exp\left\{p\log\frac{y}{K-y} - \frac{2a}{\sigma^2}\frac{K-y}{y} - \frac{2\mu}{\sigma^2}\frac{y}{K-y} + 2\log\left(\frac{K}{K-y}\right)\right\} \\ &\quad \times\left|-\frac{1}{\sigma}\left(\frac{1}{y}+\frac{1}{K-y}\right)\right| \\ &= \mathcal{N}^{-1}\frac{1}{y(K-y)^3}\left(\frac{y}{K-y}\right)^p \exp\left\{-\frac{2a}{\sigma^2}\frac{K-y}{y} - \frac{2\mu}{\sigma^2}\frac{y}{K-y}\right\}, \quad (1.20) \end{aligned}$$

where

$$p = -2D/\sigma = \frac{2}{\sigma^2}(a + b - \mu) - 1, \tag{1.21}$$

and

$$\mathcal{N} = \int_0^K \frac{1}{y(K-y)^3} \left(\frac{y}{K-y}\right)^p \exp\left\{-\frac{2a}{\sigma^2}\frac{K-y}{y} - \frac{2\mu}{\sigma^2}\frac{y}{K-y}\right\} dy. \tag{1.22}$$

We notice that, in the considered case, the integral in (1.22) is finite. By considering the change of variable:

$$u = \frac{y}{K-y}, \tag{1.23}$$

the normalization constant can be expressed as:

$$\begin{aligned}
\mathcal{N} &= \frac{1}{K^3}\int_0^{+\infty} u^{p-1}\,(1+u)^2\, e^{-(\alpha u+\beta/u)/2} du \\
&= \frac{1}{K^3}\int_0^{+\infty} u^{p+1} e^{-(\alpha u+\beta/u)/2} du \\
&+ \frac{2}{K^3}\int_0^{+\infty} u^{p} e^{-(\alpha u+\beta/u)/2} du \\
&+ \frac{1}{K^3}\int_0^{+\infty} u^{p-1} e^{-(\alpha u+\beta/u)/2} du,
\end{aligned}$$

where $\alpha = 4\mu/\sigma^2$, $\beta = 4a/\sigma^2$. We notice that, the integrals in the previous formulas can be expressed in terms of appropriate special functions. In fact, remembering that the integral representation of the modified Bessel function of second kind $K_q(\cdot)$ of index q is:

$$\begin{aligned}
K_q(x) &= \frac{1}{2}\int_0^{+\infty} \exp(-x\cosh u)\cosh(qu) du \\
&= \int_0^{+\infty} \exp\left\{-\frac{x}{2}(u + u^{-1})\right\} u^{q-1} du,
\end{aligned}$$

we obtain:

$$\mathcal{N} = \frac{1}{K^3}\left(\frac{\beta}{\alpha}\right)^{p/2}\left\{K_p(\sqrt{\alpha\beta}) + 2\left(\frac{\beta}{\alpha}\right)^{1/2}K_{p+1}(\sqrt{\alpha\beta}) + \left(\frac{\beta}{\alpha}\right)K_{p+2}(\sqrt{\alpha\beta})\right\}.$$

Similar calculations allow us to express the first two moments m, m_2 of the r.v. distributed as (1.20) in terms of the modified Bessel functions of the second kind:

$$\begin{aligned} m &= \frac{1}{\mathcal{N}K^2}\int_0^{+\infty} u^p\,(1+u)\,e^{-(\alpha u+\beta/u)/2}du \\ &= K\frac{\sqrt{\beta/\alpha}K_{p+1}(\sqrt{\alpha\beta}) + (\beta/\alpha)K_{p+2}(\sqrt{\alpha\beta})}{2\left[\sqrt{\beta/\alpha} - (p+1)/\sqrt{\alpha\beta}\right]K_{p+1}(\sqrt{\alpha\beta}) + [1+(\beta/\alpha)]\,K_{p+2}(\sqrt{\alpha\beta})}; \\ m_2 &= \frac{1}{\mathcal{N}K}\int_0^{+\infty} u^{p+1}e^{-(\alpha u+\beta/u)/2}du \\ &= \frac{1}{\mathcal{N}K}\,(\alpha/\beta)^{-p/2-1}\,K_{p+2}(\sqrt{\alpha\beta}), \end{aligned}$$

where we have used the recursive relation:

$$K_p(x) = K_{p+2}(x) - \frac{2(p+1)}{x}K_{p+1}(x).$$

It is intersting to study the deterministic limit ($\sigma^2 \to 0$) of the above expressions. Direct analysis is difficult because both the indices and the arguments of the K-functions depend on parameter σ^2. However, asymptotic behaviour can be analyzed using Laplace's approximation method. In fact, a simple re-writing of $\mathcal{N}, m$ and m_2 provides:

$$
\begin{aligned}
\mathcal{N} &= \frac{1}{K^3}\int_0^{+\infty} u^{-2}(1+u)^2 \exp\left\{\frac{2}{\sigma^2}\left[(b+a-\mu)\log u - \mu u - \frac{a}{u}\right]\right\} du, \\
m &= \frac{1}{\mathcal{N}K^2}\int_0^{+\infty} u^{-1}(1+u) \exp\left\{\frac{2}{\sigma^2}\left[(b+a-\mu)\log u - \mu u - \frac{a}{u}\right]\right\} du, \\
m_2 &= \frac{1}{\mathcal{N}K}\int_0^{+\infty} \exp\left\{\frac{2}{\sigma^2}\left[(b+a-\mu)\log u - \mu u - \frac{a}{u}\right]\right\} du.
\end{aligned}
$$

Thus, if we let

$$
u^* = \arg\max_{u\in\mathbb{R}^+} \{g(u)\},
$$

where

$$
g(u) = (b+a-\mu)\log u - \mu u - \frac{a}{u}, \quad u > 0,
$$

we can use the following approximation, valid for small σ^2:

$$
\int_0^{+\infty} h(u) e^{2g(u)/\sigma^2} du \approx \sqrt{\frac{\pi\sigma^2}{|g''(u^*)|}} h(u^*) e^{2g(u^*)/\sigma^2}.
$$

Using this approximation for the integrals appearing in the definitions of $\mathcal{N}, m$, and m_2, we obtain:

$$
\begin{aligned}
m &\approx \frac{Ku^*}{1+u^*} \\
m_2 &\approx m^2 = \left(\frac{Ku^*}{1+u^*}\right)^2.
\end{aligned}
$$

Straightforward calculations show that

$$
u^* = \frac{b+a-\mu+\sqrt{(b+a-\mu)^2+4a\mu}}{2\mu},
$$

and

$$m^* \doteq K\frac{u^*}{1+u^*} = K\frac{b-a-\mu+\sqrt{(b-a-\mu)^2+4ab}}{2b}. \tag{1.24}$$

We conclude that

$$\lim_{\sigma^2\to 0} m = m^*, \qquad \lim_{\sigma^2\to 0}(m_2 - m^2) = 0.$$

Note that the limit value (1.24) concides with the positive solution y^+ in (1.2). In other words, as one could expect, the equilibrium distribution density of the SBM in the small noise limit tends to be concentrated around the (positive) stationary point of the corresponding deterministic equation.

As an important remark, we notice that the stationary density (1.20) does depend on the volatility parameter σ only through the ratios $a/\sigma^2, b/\sigma^2, \mu/\sigma^2$. In other words, the stationary distribution density is invariant with respect to the transformations $\sigma^2 \to c\,\sigma^2$, $a \to c^2\,a$, $b \to c^2\,b$, $\mu \to c^2\,\mu$ with c arbitrary positive constant. This implies that the model parameters are not identifiable. Lack of parameter identifiability for the stationary distribution has important consequences in parametric inference for the SBM (see Chapter 4).

ii) $\mathbf{a = 0}$, $\mu > \mathbf{0}$ (no self-innovation, but possibility of recessing).

We first consider the case where:

$$\sigma^2 < 2(b-\mu).$$

In this case:

$$\begin{aligned} &\lim_{x\to\pm\infty} A(x) \\ = &\lim_{x\to\pm\infty}\left[\left(\frac{\mu}{\sigma}-\frac{b}{\sigma}+\frac{\sigma}{2}\right)x - \frac{\mu}{\sigma^2}e^{-\sigma x} + \log(1+e^{-\sigma x}) - K_2\right] = -\infty. \end{aligned}$$

As in the previous case, the positivity of $q(\cdot)$ implies $C = 0$, so that:

$$q(x) = K_1 e^{2A(x)} = \mathcal{N}_0^{-1} \exp\left\{2\left(\frac{\mu}{\sigma} - \frac{b}{\sigma} + \frac{\sigma}{2}\right)x - 2\frac{\mu}{\sigma^2}e^{-\sigma x} + 2\log(1+e^{-\sigma x})\right\} \tag{1.25}$$

where $\mathcal{N}_0$ is to be determined by normalization. Returning to the original variable, the equilibrium density is:

$$p(y) = \mathcal{N}_0^{-1} \frac{1}{y(K-y)^3}\left(\frac{y}{K-y}\right)^p \exp\left\{-\frac{2\mu}{\sigma^2}\frac{y}{K-y}\right\}, \tag{1.26}$$

where

$$\mathcal{N}_0 = \int_0^K \frac{1}{y(K-y)^3}\left(\frac{y}{K-y}\right)^p \exp\left\{-\frac{2\mu}{\sigma^2}\frac{y}{K-y}\right\} dy, \tag{1.27}$$

and $p = 2b/\sigma^2 - 2\mu/\sigma^2 - 1 > 0$. The change of variable $z = \frac{2\mu}{\sigma^2}y/(K-y)$ in (1.27) provides:

$$\mathcal{N}_0 = \frac{1}{K^3}\left(\frac{\sigma^2}{2\mu}\right)^p \int_0^{+\infty}\left(1+\frac{\sigma^2}{2\mu}z\right)^2 z^{p-1}e^{-z}dz.$$

Expliciting the squared term in the integrand in the previous expression, we obtain:

$$\begin{aligned}\mathcal{N}_0 &= \frac{1}{K^3}\left(\frac{\sigma^2}{2\mu}\right)^p \Gamma(p)\left[1+\frac{\sigma^2}{\mu}p + \frac{\sigma^4}{4\mu^2}p(p+1)\right] \\ &= \frac{1}{K^3}\left(\frac{\sigma^2}{2\mu}\right)^p \Gamma(p)\frac{1}{\mu^2}\left[b^2 - \frac{\sigma^2}{2}(\mu+b)\right],\end{aligned}$$

where we have used the relation $\Gamma(\xi+1) = \xi\Gamma(\xi)$, with $\Gamma(\cdot)$ denoting the Gamma function. The expected value m of the r.v. whose distribution density is $p(\cdot)$, is given by:

$$m = \mathcal{N}_0^{-1}\int_0^K \frac{1}{(K-y)^3}\left(\frac{y}{K-y}\right)^p \exp\left\{-\frac{2\mu}{\sigma^2}\frac{y}{K-y}\right\} dy. \tag{1.28}$$

Expressing again the integral in terms of Gamma functions and using the expression of $\mathcal{N}_0$, it can be proved that:

$$m = K\frac{b\,(b-\mu-\sigma^2/2)}{b^2-\sigma^2\,(\mu+b)\,/2}. \tag{1.29}$$

Similar calculations lead to the following expression for the second moment:

$$m_2 = K^2(b-\mu)\left(b-\mu-\frac{\sigma^2}{2}\right)\left(b^2-\frac{\sigma^2}{2}\,(\mu+b)\right)^{-1}. \tag{1.30}$$

Then, it follows that:

$$m_2 - m^2 = \frac{(K\mu\sigma)^2(b-\mu-\frac{\sigma^2}{2})}{2\left[b^2-\frac{\sigma^2}{2}(b+\mu)\right]^2}.$$

Remark. It is interesting to study the behaviour of the first two moments and in the "deterministic limit", i.e. as $\sigma^2 \to 0$. It is easily seen that:

$$\lim_{\sigma^2\to 0} m = K\frac{b-\mu}{b}; \quad \lim_{\sigma^2\to 0}(m_2-m^2) = 0. \tag{1.31}$$

Equalities (1.31) show that, as one would expect, the equilbrium distribution tends to become singular with mass on the deterministic stationary point, as the noise vanishes. Furthermore, it is easily checked that, for arbitrary values of σ^2

$$\lim_{\mu\to 0} m = K; \qquad \lim_{\mu\to 0}(m_2-m^2) = 0. \tag{1.32}$$

The last equations imply that the equilibrium distribution tends to become concentrated around the point $y = K$ as the mortality parameter vanishes.

In the special case:

$$\sigma^2 = 2(b - \mu),$$

$A(x)$ tends to a constant as $x \to +\infty$, so that $q(x)$ cannot vanish at $x = +\infty$.

Then, in this case the equilibrium distribution density does not exist.

Now we deal with the case: $\sigma^2 > 2(b - \mu)$.
In this case,

$$\begin{aligned} & \lim_{x\to+\infty} A(x) \\ = & \lim_{x\to+\infty} \left[\left(\frac{\mu}{\sigma} - \frac{b}{\sigma} + \frac{\sigma}{2} \right) x - \frac{\mu}{\sigma^2} e^{-\sigma x} + \log(1 + e^{-\sigma x}) - K_2 \right] = +\infty. \end{aligned}$$

If $C = 0$, then

$$\lim_{x\to+\infty} q(x) = \lim_{x\to+\infty} K_1 e^{2A(x)} = +\infty.$$

In the opposite case, by applying the L'Hopital's rule we find:

$$\begin{aligned} & \lim_{x\to+\infty} q(x) \\ = & \lim_{x\to+\infty} e^{2A(x)} \left[K_1 + 2C \int_d^x e^{-2A(t)} dt \right] = C \left(-\frac{\mu}{\sigma} + \frac{b}{\sigma} - \frac{\sigma}{2} \right)^{-1} \neq 0. \end{aligned}$$

Therefore, also in this case the equilibrium distribution density does not exist.
Since in the case of no self innovation ($a = 0$), both drift and diffusion functions vanish for $Y = 0$, the process $Y_t \equiv 0$ is a (trivial) solution of (1.14). Thus, it is interesting to study the stability properties of this trivial solution when the parameter are such that an invariant distribution does not exist. In particular, in analogy with the deterministic framework, it is of interest to analyze the qualitative behaviour of solutions corresponding to initial conditions y_0 that belong to a small right interval of 0. The analysis of the stochastic stability is based on the Lyapunov theory (see Appendix C for definition of stochastic stability and Lyapunov function). Without loss

of generality, we set $K = 1$.

Theorem 1.2.4 *The solution $Y_t \equiv 0$ of the SDE:*

$$dY_t = [bY_t(1 - Y_t) - \mu Y_t]\, dt + \sigma Y_t(1 - Y_t)dW_t \tag{1.33}$$

is stochastically asymptotically stable for $\sigma^2 > 2(b - \mu)$.

Proof.

The generator of SDE (1.33) is:

$$\mathcal{L} = [by(1 - y) - \mu y]\, \frac{\partial}{\partial y} + \frac{\sigma^2}{2} y^2 (1 - y)^2 \frac{\partial^2}{\partial y^2}.$$

Let us define the Lyapunov function:

$$V(y) = \left| \int_0^y g(x)dx \right|,$$

where

$$g(y) \doteq \exp\left\{ -A \log\left| \frac{y}{1 - y} \right| + \frac{2b}{\sigma^2} \frac{y}{1 - y} \right\}.$$

Assuming $A < 1$, so that V is well defined in a neighbourood of $\{0\}$, we have:

$$\begin{aligned} \mathcal{L}V(y) &= \operatorname{sgn}(y) g(y) \left[by(1 - y) - \mu y \right] \\ &\quad + \operatorname{sgn}(y) g(y) \left[-\frac{A}{y(1 - y)} + \frac{2b}{\sigma^2} \frac{1}{(1 - y)^2} \right] \frac{\sigma^2}{2} y^2 (1 - y)^2 \\ &= g(y) y \operatorname{sgn}(y) \left[b - \mu - \frac{\sigma^2}{2} A + \frac{\sigma^2}{2} Ay \right], \end{aligned}$$

Since $g(y)y \operatorname{sgn}(y) > 0$, the thesis is proved if there is a neighbourood D of $y = \{0\}$, such that

$$b - \mu - \frac{\sigma^2}{2} A + \frac{\sigma^2}{2} Ay < 0 \tag{1.34}$$

for $y \in D$. In fact, this follows from the r.h.s. of (1.34) being a continuous function of (A, y), taking negative value at point $(1, 0)$.

Remark.

It is interesting to note that the presence of sufficiently strong noise can determine disadoption of all individuals with high probability, also in situations where the corresponding deterministic system goes away from zero. Specifically, this happens when $b - \mu - \sigma^2/2 \leq 0 < b - \mu$. The stationary point $y \equiv 0$ is deterministically unstable, but the solution $Y_t \equiv 0$ of the Bass SDE is asymptotically stable in probability. This situation is typical of logistic models and in population dynamics is responsible for the fact that "noise can cause extintion".

(iii) $\mu = \mathbf{0}$ (self-innovation is present but recess is not allowed).
In this case we have:

$$\begin{aligned} &\lim_{x \to -\infty} A(x) \\ = &\lim_{x \to -\infty} \left[\left(-\frac{a}{\sigma} - \frac{b}{\sigma} + \frac{\sigma}{2} \right) x - \frac{a}{\sigma^2} e^{\sigma x} + \log(1 + e^{-\sigma x}) - K_2 \right] = +\infty. \end{aligned}$$

The same analysis as in the case $\sigma^2 = 2(b - \mu)$ leads to the conclusion that the equilibrium distribution density does not exist.
Without loss of generality, let us assume that $K = 1$. We note that in treated case of no recess ($\mu = 0$), $Y_t \equiv 1$ is a solution of the equation

$$dY_t = (a + bY_t)(1 - Y_t)dt + \sigma Y_t(1 - Y_t)dW_t. \tag{1.35}$$

Thus, since the process does not admit a proper equilibrium distribution, one could ask whether $Y_t \equiv 1$ is stable in probability. Note that the transformed process defined by $Z_t = 1 - Y_t$ is solution of the SDE:

$$dZ_t = -\left[aZ_t + bZ_t(1 - Z_t)\right] dt - \sigma Z_t(1 - Z_t)dW_t, \tag{1.36}$$

with $Z_0 = 1 - Y_0$. Thus, stochastic stability of $Y_t \equiv 1$ is equivalent to stochastic stability of $Z_t \equiv 0$.

It can be easily shown that the Lyapunov function:

$$V(z) = \left| \int_0^z \exp\left\{ -\frac{2b}{\sigma^2} \frac{x}{1-x} \right\} dx \right|,$$

satisfies:

$$\mathcal{L}V(z) = -(a+b)z \operatorname{sgn}(z) \exp\left\{ -\frac{2b}{\sigma^2} \frac{z}{1-z} \right\}.$$

Thus, $\mathcal{L}V(z) < 0, \forall z \in \mathbb{R}$. This shows that the solution $Z_t \equiv 0$ of (1.36) (correspondingly, the solution $Y_t \equiv 1$ of equation (1.35)) is stable in probability.

Chapter 2

Exact Simulation

2.1 Simulation of diffusion processes

In many cases, when analytic expression for the transition density of a diffusion process is not available, one uses simulation of sample paths in order to approximate some characteristics of the process (such as expected value of the process at some fixed times, or hitting time at the boundary of some region of interest) via Monte Carlo average. One of the simplest method for simulation of solutions of SDEs is based on the Euler-Maruyama scheme (Maruyama, 1955; Kloeden and Platen, 1995). According to this scheme, given the SDE on $[0, T]$:

$$dY_t = b(t, Y_t)dt + \sigma(t, Y_t)dW_t, \quad Y(0) = y_0, \tag{2.1}$$

sample paths of the (strong) solution (provided that it exists and it is unique) are obtained by partitioning the interval $[0, T]$ in n sub-intervals $[t_{i-1}, t_i], i = 1, \ldots, n$, with $t_0 = 0$, $t_n = T$, and using a Gaussian approximation on each sub-interval. More specifically, denoting by Δt_i the time increment $t_i - t_{i-1}$ and conditionally on the value of the process at time t_i, the process increments $\Delta Y_i \doteq Y_{t_{i+1}} - Y_{t_i}$ are approximated, for $i = 0, \ldots, n-1$, by the random variables:

$$b(t_i, Y_{t_i})\Delta t_i + \sigma(t_i, Y_{t_i})\sqrt{\Delta t_i}Z_i, \tag{2.2}$$

where $Z_i \sim N(0, 1)$ are independent standard Gaussian variables. It is well known, (Kloeden and Platen, 1995; Pascucci, 2008), that under some regularity conditions, scheme (2.2) defines an "approximate process" that in the

limit $\max_i |\Delta t_i| \to 0$ converges, in a suitable metrics, to the solution of (2.1). In order to avoid possible bias deriving from Euler approximation, a very fine partition of the interval $[0, T]$ is generally taken, resulting in a very large number of Gaussian random variables to be generated. Besides computational complexity, Euler scheme has another disadvantage in case some qualitative properties have to be preserved by the simulated trajectories. For instance, if it is known that sample paths of a process Y cannot exit from some domain D, this property, reffered to as *regularity* on D, is not guaranteed by the Euler approximation. As we will see in Section 1.2, this is the case for solutions of certain classes of logistic-type stochastic differential equations that are known to have particular regularity properties. This can be a serious problem when the interest is focused on hitting time of some regions inside the domain D. In some cases it is possible to use numerical schemes to approximate solutions of SDEs that preserve the regularity properties of the solutions mantaining, at the same time, the convergence order of standard Euler scheme (Shurz, 2007). In this section we describe an approach recently introduced by Beskos, Papaspilioupolos and Roberts (2006), that allows to simulate *exact* solutions of certain classes of SDEs. The methodology is based on a suitable acceptance-rejection algorithm (AR) that provides realizations from arbitrary finite dimensional distributions of the target process. According to the terminology of Beskos, Papaspilioupolos and Roberts (2006), the algorithm will be reffered to as *exact algorithm* (EA). A major advantage with this approach is that, as we will see in the following, a slightly different version of the EA allows one to draw realizations from diffusion processes *conditioned* on the ending point of a (possibly d-dimensional) interval (diffusion bridges). Simulation of diffusion bridges is often necessary in many techniques for inference on diffusion processes, but it is difficult to implement with traditional numerical schemes.

We start by illustrating the methodology in the time-homogeneous one-dimensional case.

2.2 Exact Algorithm

2.2.1 Approach based on change of measure

Let $Y_t,\ t \in [0, T]$ be a strong solution of the SDE:

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \tag{2.3}$$

with initial condition $Y(0) = y_0 \in \mathbb{R}$. Note that both the drift and the diffusion coefficients depend on time only through the process (homogeneity).

We want to obtain a realization of the n-dimensional random variable $(Y_{t_1}, \ldots, Y_{t_n})$, where $0 \leq t_1 < \ldots < t_n \leq T$ are arbitrary time instants. We will first draw the process at a finite collection of random times $\tau_1 < \tau_2 < \ldots < \tau_{\mathcal{K}}$ ($\mathcal{K}$ random) in $[0, T]$, and then we will obtain Y_{t_i}, $(i = 1, \ldots, n)$ exploiting the Markov property of the process Y and using the dynamic of a suitable "bridge" process. We can assume, without loss of generality, $t_1 = 0$, $t_n = T$. The first step is to transform the process Y in a new process X verifying a SDE with unit diffusion coefficient. This can be easily done through the transformation $X_t = \eta(Y_t)$ where:

$$\eta(y) \doteq \int_c^y \frac{1}{\sigma(u)} du, \tag{2.4}$$

with c arbitrary constant, provided that the integral above does exist.

Transformation (2.4) is sometimes referred to as Lamperti transform and is frequent in applications to inference problems where change of measure is used (Florens, 1999; Ait-Sahalia, 2002).

Applying Ito's lemma to the transformation (2.4), it is easily seen that the new process X verifies the SDE:

$$dX_t = \alpha(X_t)dt + dW_t, \qquad X(0) = x_0 = \eta(y_0), \tag{2.5}$$

with

$$\alpha(x) = \frac{b\left[\eta^{-1}(x)\right]}{\sigma[\eta^{-1}(x)]} - \frac{1}{2}\sigma'\left[\eta^{-1}(x)\right]. \tag{2.6}$$

Since realizations from the law of Y can be obtained from realizations from the law of X, by simply inverting the trasformation (2.4), hereafter we will

deal with the problem of drawing from the law of a diffusion process X satisfying equation (2.5).
The core of the methodology is the Girsanov theorem on change of measure (see, for instance, Karatzas and Shreve (1991)). According to this theorem, given the canonical set-up $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ the process X, the solution of the equation (2.5) on $[0, T]$, is a Brownian motion with respect to the measure Q defined by:

$$Q(A) = E_P(Z\mathbf{1}_A),$$

where E_P denotes expectation with respect to the natural measure P, $A \in \mathcal{F}_T$, and Z is the exponential martingale:

$$Z_T = \exp\left\{-\int_0^T \alpha(X_s)dW_s - \frac{1}{2}\int_0^T \alpha(X_s)^2 ds\right\}. \tag{2.7}$$

Remark. Actually, the process (2.7) is guaranteed only to be a *local* martingale. In order to conclude that X_t is a Brownian motion w.r.t. Q, one should prove that Z is in fact a martingale.

An important consequence of the Girsanov theorem is that it implies that the distribution laws of diffusion processes having the same diffusion coefficient are absolutely continuous with respect to each other. In particular, let $\mathbb{P}^{x_0}$ be the probability measure induced by the process X, solution of the SDE (2.5), on the space $(\mathcal{C}[0, T], \mathcal{B})$ of the continuous functions on $[0, T]$, equipped with the σ-algebra of the cylinder sets $\mathcal{B}$. Then, the Radon-Nikodym derivative of $\mathbb{P}^{x_0}$ with respect to the Wiener measure $\mathbb{W}^{x_0}$, corresponding to the Brownian motion starting in x_0, can be expressed as:

$$\frac{d\mathbb{P}^{x_0}}{d\mathbb{W}^{x_0}} = \exp\left\{\int_0^T \alpha(W_s)dW_s - \frac{1}{2}\int_0^T \alpha(W_s)^2 ds\right\}. \tag{2.8}$$

Formula (2.8) can be simplified by eliminating the stochastic integral appearing on the right hand side. To this aim, define:

$$A(x) = \int_a^x \alpha(u)du, \tag{2.9}$$

where a is an arbitrary constant, and introduce the new process:

$$U_t = A(W_t).$$

Ito's formula provides:

$$U_t = U_0 + \int_0^t \frac{1}{2}\alpha'(W_s)ds + \int_0^t \alpha(W_s)dW_s,$$

thus we obtain:

$$\frac{d\mathbb{P}^{x_0}}{d\mathbb{W}^{x_0}} = \exp\left\{A(W_T) - A(x_0) - \frac{1}{2}\int_0^T \left[\alpha(W_s)^2 + \alpha'(W_s)\right]ds\right\}. \quad (2.10)$$

In formula (2.10), and in the rest of this subsection, $W_\cdot$ is to be intended as the typical element of $\mathcal{C}[0,T]$ rather than a Brownian motion.

Remark. Note that, although the Girsanov theorem remains valid if W is a d dimensional Brownian motion for $d \geq 1$, however for $d > 1$, the elimination of the stochastic integral in formula (2.8) is possible only if the function $\alpha(\cdot)$ verifies some extra conditions (see Section 2.2.7).

Now, following Beskos, Papaspilioupolos and Roberts (2006), we introduce the *biased Brownian motion* defined as the stochastic process $W_{h_T}^{x_0}$ whose law is that of the Brownian motion starting from x_0 and conditioned on having marginal density $h_T(\cdot)$ at the end of time interval T, where $h_T(u)$ is proportional to $\exp\{A(u) - (u-x_0)^2/2T\}$, $u \in \mathbb{R}$ (we assume that h_T is integrable). Let $\mathbb{Z}_T^{x_0}$ be the measure on $(\mathcal{C}[0,T], \mathcal{B})$ associated with the process $W_{h_T}^{x_0}$. Then, a change of measure argument shows that:

$$\frac{d\mathbb{P}^{x_0}}{d\mathbb{Z}_T^{x_0}} \propto \exp\left\{-\frac{1}{2}\int_0^T \left[\alpha(W_s)^2 + \alpha'(W_s)\right]ds\right\}. \quad (2.11)$$

Formula (2.11) can be justified using the (formal) relation:

$$\frac{d\mathbb{P}^{x_0}}{d\mathbb{Z}_T^{x_0}} = \frac{d\mathbb{P}^{x_0}}{d\mathbb{W}^{x_0}}\frac{d\mathbb{W}^{x_0}}{d\mathbb{Z}_T^{x_0}} \quad (2.12)$$

and the following proposition (Beskos, Papaspilioupolos and Roberts, 2006):

Proposition 1 *Let M, N be two stochastic processes defined on $[0,T]$ and $\mathbb{M}$, $\mathbb{N}$ the corresponding measures on $(\mathcal{C}[0,T], \mathcal{B})$. Let f_M, f_N be the density functions of the random variables M_T, N_T respectively, corresponding to the value of the processes at the ending point T and assume that f_M and f_N have the same support. Then, if for all $y \in \mathbb{R}$, the processes obtained by conditioning M and N on the events $\{M_T = y\}$, $\{N_T = y\}$ respectively have the same law, it holds:*

$$\frac{d\mathbb{M}}{d\mathbb{N}} = \frac{f_M}{f_N}.$$

Proof As usually, we denote by $W_\cdot$ the typical element of $(\mathcal{C}[0,T], \mathcal{B})$. The statement in the proposition about equality in law of the conditioned processes can be formally expressed as:

$$\mathbb{M}(A|\sigma(W_T)) = \mathbb{N}(A|\sigma(W_T)) \quad \mathbb{N}\ a.s., \quad A \in \mathcal{B},$$

where $\sigma(W_T)$ denotes the σ-algebra generated by the r.v. W_T. Thus we must prove that, for all $A \in \mathcal{B}$:

$$\mathbb{M}(A) = E_{\mathbb{N}}\left[\mathbf{1}_A \frac{f_M}{f_N}(W_T)\right].$$

Using the chain property of expectations:

$$\begin{aligned} E_{\mathbb{N}}\left[\mathbf{1}_A \frac{f_M}{f_N}(W_T)\right] &= E_{\mathbb{N}}\left[E_{\mathbb{N}}\left[\mathbf{1}_A \frac{f_M}{f_N}(W_T)|\sigma(W_T)\right]\right] \\ &= E_{\mathbb{N}}\left[\frac{f_M}{f_N}(W_T) \cdot E_{\mathbb{N}}\left[\mathbf{1}_A|\sigma(W_T)\right]\right] = E_{\mathbb{N}}\left[\frac{f_M}{f_N}(W_T) \cdot \mathbb{N}\left[A|\sigma(W_T)\right]\right], \end{aligned}$$

where the third equality follows from $\frac{f_M}{f_N}(W_T)$ being $\sigma(W_T)$- measurable.

The last term is the $\mathbb{N}$-expectation of a $\sigma(W_T)$-measurable r.v. and thus it can be expressed as expectation with respect to the restriction of $\mathbb{N}$ to $\sigma(W_T)$, i.e., the density f_N. Thus

$$
\begin{aligned}
E_{\mathbb{N}}\left[\frac{f_M}{f_N}(W_T)\cdot \mathbb{N}\left[A|\sigma(W_T)\right]\right] &= E_{f_N}\left[\frac{f_M}{f_N}(W_T)\cdot \mathbb{N}\left[A|\sigma(W_T)\right]\right] \\
&= E_{f_M}\left[\mathbb{N}\left[A|\sigma(W_T)\right]\right] = E_{f_M}\left[\mathbb{M}\left[A|\sigma(W_T)\right]\right] \\
&= E_{\mathbb{M}}\left[\mathbb{M}\left[A|\sigma(W_T)\right]\right] = E_{\mathbb{M}}\left[A\right]. \quad \diamondsuit
\end{aligned}
$$

If in proposition (1) we take $\mathbb{M} = \mathbb{Z}_T^{x_0}$, $\mathbb{N} = \mathbb{W}^{x_0}$, we obtain:

$$
\frac{d\mathbb{W}^{x_0}}{d\mathbb{Z}_T^{x_0}}(W) = \frac{\mathcal{N}(x_0;T)}{h_T}(W_T) \ \propto \ e^{-A(W_T)},
$$

where $\mathcal{N}(\mu;\sigma^2)$ is the Gaussian density with mean μ and variance σ^2. Using the last expression in (2.12) and taking into account (2.10), we obtain (2.11).

2.2.2 Retrospective Sampling

Formula (2.11) expresses the Radon-Nykodyn derivative of the target measure associated to the process X in terms of an ordinary integral (although the integrand is random). Under some additional conditions, it can be used to implement an acceptance-rejection algorithm for exact simulation of strong solutions of SDEs. Details are provided below.

Let $\alpha(\cdot)$ be the drift in the standardized equation (2.5) and assume that the function $\alpha(\cdot)^2+\alpha'(\cdot)$ is bounded below. Define the non negative function:

$$
\Phi(u) \doteq \frac{\alpha(u)^2+\alpha'(u)}{2} - l, \tag{2.13}
$$

where

$$
l \leq \inf_{u\in\mathbb{R}} \frac{\alpha(u)^2+\alpha'(u)}{2}.
$$

We have:

$$
\frac{d\mathbb{P}^{x_0}}{d\mathbb{Z}_T^{x_0}} \propto \exp\left\{-\int_0^T \Phi(W_s)ds\right\} \leq 1. \tag{2.14}
$$

On the basis of expression (2.14) a rejecton sampling scheme can be adopted to draw sample paths from $\mathbb{P}^{x_0}$. In fact, an ideal (impossible) procedure would be that of drawing a path from the dominating measure $\mathbb{Z}_T^{x_0}$ and accepting it with probability:

$$P_a(W_\cdot) = \exp\left\{-\int_0^T \Phi(W_s)ds\right\}. \tag{2.15}$$

The previous scheme is impossible in the sense that it would require random drawing a continuous path from $\mathbb{Z}^{x_0}$ and evaluating the integral (2.15). Of course, one can use a discrete approximation of (2.15), and simulate from the finite dimensional distribution of the process $W_{h_T}^{x_0}$ corresponding to the partition used for the discretization. However, for a certain class of diffusion processes, *exact* simulation from $\mathbb{P}^{x_0}$ can be performed by *retrospective* sampling. This is based on noting that, *given* an element of $\mathcal{C}[0,T]$, the acceptance probability (2.15) can be recognized as the probability of a certain event associated with a suitable Poisson process. Specifically, let Ψ be the homogeneous Poisson process with unit intensity on $[0,T]\times\mathbb{R}$, and N the number of points under the graph of the function $\{t \to \Phi(W_t)\}$. Then, $P_a(W_\cdot) = P\{N = 0 \mid W_t; t \in [0,T]\}$, i.e., the acceptance probability can be viewed as the probability that, conditionally on W_t, $t \in [0,T]$, no point is below the graph of $\Phi(W_\cdot)$ on $[0,T]$. The idea of restrospective sampling is to invert the usual order for drawing the proposal and the "decision variable" in acceptance-rejection (A/R) scheme (see Appendix A). This allows to implement an A/R algorithm such that the "proposal" is accepted with the right probability without knowing the proposed path on the entire interval $[0,T]$. More in detail, suppose we are able to draw $\mathcal{K}$ random points ($\mathcal{K}$ random) from Ψ on a domain in $\mathbb{R}^2$, which contains with certainty the region $\{(t,\psi)\in\mathbb{R}^2 \mid 0 \le t \le T; 0 \le \psi \le \Phi(W_t)\}$. Let these points be $\{\Psi_1,\ldots,\Psi_{\mathcal{K}}\} = \{(\tau_1,\psi_1),\ldots,(\tau_{\mathcal{K}},\psi_{\mathcal{K}})\}$. Then, we could draw a vector $x_1,\ldots,x_{\mathcal{K}}$ from the $\mathcal{K}$-dimensional marginal distribution corresponding to the times $\tau_1,\ldots,\tau_{\mathcal{K}}$ of the process $W_{h_T}^{x_0}$ and check the $\mathcal{K}$ conditions: $\Phi(x_i) \le \psi_i, \quad i = 1,\ldots,\mathcal{K}$ (see figure 2.1). If all these conditions are satisfied, no points from the Poisson process are below the graph of $\Phi(W_\cdot)$. In other words, using the previous notation, one can check the condition $N = 0$, checking only a *finite* number of conditions involving the random points from the Poisson process.

If the condition $N = 0$ holds, the *whole trajectory* $\{W_t;\ 0 \le t \le T\}$ is accepted (see Fig.2.1). In practice, the path between each pair of consecutive points (τ_i, x_i) can be reconstructed by means of the Brownian Bridge dynam-

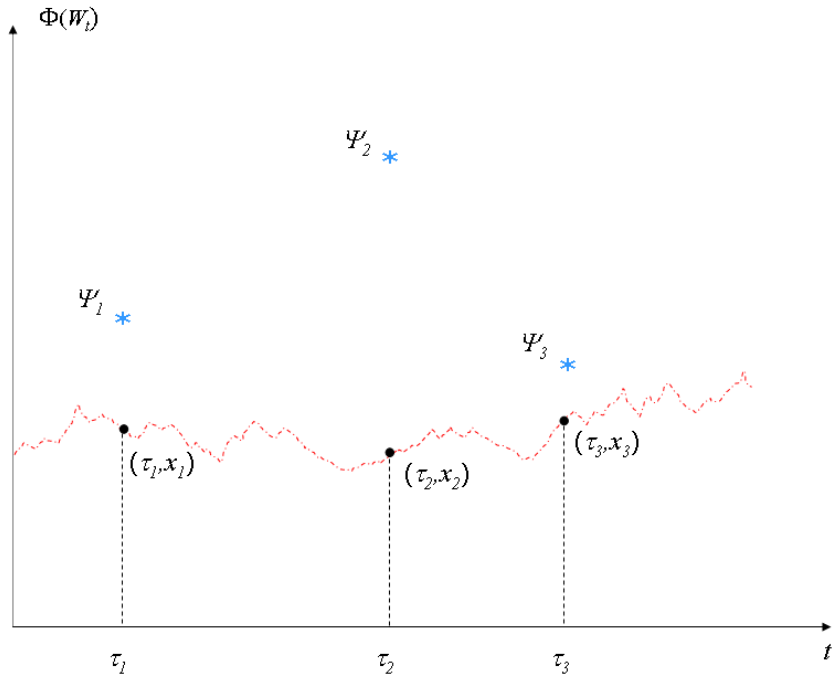

Figure 2.1: Sample path accepted by A/R algorithm. In this example $\mathcal{K} = 3$.

ics (see Appendix B for definition and properties of the Brownian Bridge). More precisely, let us suppose that we are interested in drawing a value of the process at time s, where $s_- < s < s_+$, and s_-, s_+, are the closest times on the left and on the right of s, respectively, where the values of the sample path, say x_-, x_+, are already available. Then, from the Markov property it follows that a realization from X_s can be obtained by drawing from a Brownian Bridge $BB_{(x_-,s_-)\to(x_+,s_+)}$ between x_- at time s_- and x_+ at time s_+. In particular, random draw from the distribution of X_s can be obtained by drawing from the Gaussian density function $\mathcal{N}(\mu(x_-, x_+, s_-, s_+);\ \Sigma(x_-, x_+, s_-, s_+))$ with mean:

$$\mu(x_-, x_+, s_-, s_+) = x_- + \frac{s - s_-}{s_+ - s_-}(x_+ - x_-)$$

and variance:

$$\Sigma(x_-, x_+, s_-, s_+) = \frac{(s - s_-)(s_+ - s)}{(s_+ - s_-)}.$$

The important point here is that the realization of the auxiliary random element used to decide if a path is to be accepted or not is available before (collection of istances of) the path itself is generated. This inversion is the essence of the "restrospective" approach and allows to perform exact simulation through the acceptance-rejection scheme, even though the knowledge of the path drawn from the dominating measure is limited to a finite (although random) collection of points.

2.2.3 EA1 Algorithm

From a practical point of view, exact algorithm requires drawing from a Poisson process on a two-dimensional region containing the graph of the function $t \to \Phi(W_t),\ \ t \in [0, T]$. By assumption, the function Φ is bounded below. If an upper bound, say M, does also exist for Φ, then it is sufficient drawing from the Poisson process Ψ in the rectangle $[0, T] \times [0, M]$. This is the simplest version of the Exact Algorithm, and it will be reffered to as EA1. A detailed description of EA1 for simulation of sample paths from the solution of SDE (2.5) is provided below.

EA1 algorithm

1. Draw the value x_T of the process X on the right extreme of the interval $[0, T]$ from the law of the r.v. with density $h_T(u)$ proportional to:

$$\exp\left\{A(u) - (u - x_0)^2/2T\right\}.$$

2. Draw a number $\mathcal{K}$ from a Poisson r.v. with parameter $\lambda = M \times T$.

3. Draw $\mathcal{K}$ points $(\tau_1, \psi_1) \ldots, (\tau_\mathcal{K}, \psi_\mathcal{K})$ from a two-dimensional uniform r.v. in the rectangular region $[0, T] \times [0, M]$.

4. At times $(\tau_1, \ldots, \tau_\mathcal{K})$, draw $\mathcal{K}$ instances $W_{\tau_1}, \ldots, W_{\tau_\mathcal{K}}$ of the Brownian Bridge $BB_{(0,x_0)\to(T,x_T)}$.

5. Compute the function Φ in $W_{\tau_1}, \ldots, W_{\tau_\mathcal{K}}$. Let $x_1 = \Phi(W_{\tau_1}), \ldots, x_\mathcal{K} = \Phi(W_{\tau_\mathcal{K}})$. If the inequalities: $x_1 \le \psi_1, \ldots, x_\mathcal{K} \le \psi_\mathcal{K}$ are verified accept the path, otherwise go to step 1.

2.2.4 EA2 Algorithm

If the function Φ defined in Section 2.2.2 is not bounded above, the exact algorithm based on retrospective sampling is more involved. In fact, a region where the graph of the function $t \to \Phi(W_t)$ lies with certainty does not exist. As a consequence, one has the practical problem of drawing points from the auxiliary Poisson process used to decide whether the proposed path has to be accepted or rejected. However, if one of the two conditions:

$$\lim_{u \to \pm\infty} \Phi(u) < +\infty \tag{2.16}$$

holds, a relatively simple version of the exact algorithm can still be used. In the following we illutrate the version described by Beskos, Papaspilioupolos and Roberts (2006) named EA2. To fix ideas, suppose that $\lim_{u\to+\infty} \Phi(u) < +\infty$. Then, it follows that for any $m \in \mathbb{R}$, we can find a number $M(m)$ such that:

$$\sup_{u \in [m,+\infty)} \Phi(u) < M(m). \tag{2.17}$$

Thus, if we knew the minimum m attained by the path on $[0, T]$, we could proceed as in algorithm EA1, and draw from a unit intensity Poisson process on $[0, T] \times [0, M(m)]$ with $M(m)$ satisfying (2.17). This suggests a sligthly different version of the exact algorithm that allows to apply the retrospective sampling approach to a much wider class of diffusion processes.

EA2 algorithm

1. Draw the value x_T of the process X on the right extreme of the interval $[0, T]$ from the r.v. with density $h_T(u)$ proportional to:

$$\exp\left\{A(u) - (u - x_0)^2/2T\right\}.$$

2. Draw the minimum m of a Brownian Bridge between x_0 at time $t = 0$ and x_T at time $t = T$, and the time t_m at which it is attained.

3. Find an upper bound $M(m)$ for the function $\Phi(\cdot)$ on $[m, +\infty)$.

4. Draw a number $\mathcal{K}$ from a Poisson r.v. with parameter $\lambda = M(m) \times T$.

5. Draw $\mathcal{K}$ points $(\tau_1, \psi_1) \ldots, (\tau_\mathcal{K}, \psi_\mathcal{K})$ from a two-dimensional uniform r.v. in the rectangular region $[0, T] \times [0, M(m)]$.

6. At times $(\tau_1, \ldots, \tau_\mathcal{K})$, draw $\mathcal{K}$ instances $W_{\tau_1}, \ldots, W_{\tau_\mathcal{K}}$ of the Brownian Bridge $BB_{(0,x_0)\to(T,x_T)}$ *conditioned on having its minimum* m *at time* t_m.

7. Compute the function Φ in $W_{\tau_1}, \ldots, W_{\tau_\mathcal{K}}$. Let $x_1 = \Phi(W_{\tau_1}), \ldots, x_\mathcal{K} = \Phi(W_{\tau_\mathcal{K}})$. If the inequalities: $x_1 \leq \psi_1, \ldots, x_\mathcal{K} \leq \psi_\mathcal{K}$ are verified accept the path, otherwise go to step 1.

The main difference between EA1 and EA2 is that in EA2 the proposed trajectories are drawn conditionally on the value of their minimum on $[0, T]$ and on the time at which the minimum is achieved. In order to implement EA2 we must be able to draw from the joint distribution of m and t_m, and to generate paths of a Brownian Bridge conditionally on the realized values of these random variables. Fortunately, both these problems can be easily solved by using the results reported below. As far as the first problem is concerned, the joint distribution of interest is given by the following theorem.

Theorem 2.2.1 *Let* $W = \{W_t; 0 \leq t \leq T\}$ *be a standard Brownian motion* $(W_0 = 0)$ *on* $[0, T]$*, and define:*

$$\begin{aligned} m_T &\doteq \inf\{W_t;\ \ 0 \leq t \leq T\} \\ \theta &\doteq \sup\{t \in [0, T] \mid W_t = m_T\}. \end{aligned}$$

Then, the pdf $f^a_{m_T,\theta}(m, t_m)$ *of the joint distribution of the random variables* m_T *and* θ*, conditional on* $\{W_T = a\}$ *is given by:*

$$f^a_{m_T,\theta}(m, t_m) = Z^{-1} \frac{m(m-a)}{\sqrt{t_m^3 (T - t_m)^3}} \exp\left\{-\frac{m^2}{2t_m} - \frac{(m-a)^2}{2(T-t_m)}\right\}, \quad m \leq \min\{0, a\} \tag{2.18}$$

where Z *is the normalization constant. Moreover, the marginal distribution w.r.t.* m_T *has density:*

$$f^a_{m_T}(m) = \frac{2}{T} \exp\left\{\frac{a^2}{2T} - \frac{(a-2m)^2}{2T}\right\}. \tag{2.19}$$

Proof (Karatzas and Shreve, 1991, Chapter 2, p. 225).

The following theorem provides a simple scheme to draw from the distribution (2.18).

Theorem 2.2.2 *Let m_T and θ be defined as in Theorem 2.2.1, E_1 an exponential random variable with unit mean, and*

$$Z_1 \doteq \frac{\sqrt{2TE_1 + a^2}}{2}.$$

Then

$$P\left\{m_T \in dm \mid W_T = a\right\} = P\left\{Z_1 \in dm\right\}. \tag{2.20}$$

Moreover, for a given realization m of m_T, let

$$c_1 = \frac{(a-m)^2}{2T}, \quad c_2 = \frac{m^2}{2T},$$

and let $X \sim IG(\sqrt{c_1/c_2}, 2c_1)$, $Y \sim IG(\sqrt{c_2/c_1}, 2c_2)$ be two independent inverse Gaussian random variables with parameters $(\sqrt{c_1/c_2}, 2c_1)$ and $(\sqrt{c_2/c_1}, 2c_2)$, respectively. Then, if V is a r.v. whose law is a mixture of the distributions of X and Y^{-1}, with mixing weight $(1+\sqrt{c_1/c_2})^{-1}$, the pdf of the joint distribution of Z_1, Z_2, with $Z_2 \doteq T/(1+V)$, is the same as the one in (2.18).

Proof

Remebering that the pdf $f(u)$, $(u \in \mathbb{R}^+)$ of an exponential r.v. with unit mean is e^{-u}, and using the change of variable $u \to \frac{\sqrt{2Tu+a^2}}{2}$, one can easily check that the pdf of Z_1 is (2.19). In order to prove the second statement of the theorem, we first note that:

$$\begin{aligned} & P\left\{\theta \in dt_m | W_T = a,\ m_T = m\right\} \\ \propto\ & \frac{1}{\sqrt{t_m^3(T-t_m)^3}} \exp\left\{-\frac{m^2}{2t_m} - \frac{(m-a)^2}{2(T-t_m)}\right\} dt_m. \end{aligned} \tag{2.21}$$

If we define the new random variable:

$$V \doteq \frac{T-\theta}{\theta}, \tag{2.22}$$

we find, after little algebra:

$$P\{V \in dv\} \propto \left(v^{-\frac{3}{2}} + v^{-\frac{1}{2}}\right) \exp\left\{-\frac{m^2}{2Y}v - \frac{(a-m)^2}{2Tv}dv\right\}. \tag{2.23}$$

The pdf of the Inverse Gaussian distribution with parameters $\mu > 0, \lambda > 0$ is:

$$\mathcal{IG}(v;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi v^3}} \exp\left[-\frac{\lambda(v-\mu)^2}{2\mu^2 v}\right], \quad v \geq 0 \tag{2.24}$$

(see, for instance, Shuster (1968)). Starting from the expression (2.24), it is easy to prove that, if X has density function $IG(x;\mu,\lambda)$, the pdf of the r.v. X^{-1} is:

$$\mathcal{IG}^{-1}(v;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi v}} \exp\left[-\frac{\lambda(1-\mu x)^2}{2\mu^2 v}\right], \quad v \geq 0. \tag{2.25}$$

Thus, if we set in (2.24)

$$\mu \equiv \mu_1 \doteq \frac{a-m}{m} = \sqrt{c_1/c_2}; \;\; \lambda \equiv \lambda_1 \doteq \frac{(a-m)^2}{T} = 2c_1$$

and in (2.25)

$$\mu \equiv \mu_2 \doteq \frac{m}{a-m} = \sqrt{c_2/c_1}; \;\; \lambda \equiv \lambda_2 \doteq \frac{m^2}{T} = 2c_2$$

we can re-write the density in (2.23) as:

$$f(v;\ \mu_1,\lambda_1,\mu_2,\lambda_2) = p \cdot \mathcal{IG}(v;\ \mu_1,\lambda_1) + (1-p) \cdot \mathcal{IG}^{-1}(v;\ \mu_2,\lambda_2), \tag{2.26}$$

where

$$p \doteq \frac{N_1}{N_1 + N_2},$$

and

$$N_1 \doteq \int_0^{+\infty} v^{-\frac{3}{2}} \exp\left[-\frac{\lambda_1(v-\mu_1)^2}{2\mu_1^2 v}\right] dv, \quad N_2 \doteq \int_0^{+\infty} v^{-\frac{1}{2}} \exp\left[-\frac{\lambda_2(v-\mu_2)^2}{2\mu_2^2 v}\right] dv.$$

Noting that $N_2/N_1 = \sqrt{\lambda_1/\lambda_2} = \sqrt{c_1/c_2}$ gives the thesis. $\diamondsuit$

On the basis of theorems 2.2.1 and 2.2.2, it is easy to draw from the joint distribution of the Brownian Bridge minimum and the time it is achieved. In fact, firstly one obtains a value m of the minimum by drawing from the distribution of a unit mean exponential r.v. and using the appropriate transformation. Then, for given m, a realization u from a uniform r.v. on $[0, 1]$ can be used to decide from which component of the mixture (2.26) a random draw v is needed: if $u < p$, $IG(\sqrt{c_1/c_2}, 2c_1)$ is to be used, otherwise the right distribution is $IG^{-1}(\sqrt{c_2/c_1}, 2c_2)$. Finally, the time t_m corresponding to m is obtained via the transformation: $t_m = T/(1+v)$. The above scheme requires generation from inverse Gaussian distribution. To this aim, a simple scheme is based on the following lemma due to Shuster (1968):

Lemma 1 *Let: $X \sim IG(\mu, \lambda)$. Then, the random variable:*

$$Y \doteq \frac{\lambda(X-\mu)^2}{\mu^2 X} \tag{2.27}$$

is distributed as the square of a normal random variable, i.e., it is chi-square with one degree of freedom (χ_1).

The proof of lemma 1 is straigthforward. With the aid of the lemma, it is easy to generate from an inverse Gaussian r.v.: it sufficies drawing from a χ_1-r.v., and then inverting (2.27) through some suitable method for the inversion of one-to-many transformations. A simple method for drawing from an inverse Gaussian distribution can be found in Devroye (1986).

Once we have a realization (m, t_m) of the minimum of the Brownian Bridge and of the time it is achieved, we must reconstruct the rest of the Brownian path conditionally on the already drawn elements. To this aim, we use a result from Asmussen et al. (1995) which provides the decomposition of the Brownian path given its minimum and the values at the extremal points of an interval (and the corresponding times) in terms of two independent Bessel processes.

Proposition 2 *Let the process:*

$$W^c \doteq \{W \mid m_T = m,\ \theta = t_m,\ W_T = a\}$$

be the Brownian motion on the interval $[0, T]$ *conditioned on achieving its minimum* m *at time* t_m *and terminating at point* a $(W_T = a)$*. Then, the two processes:*

$$\{W_s^c; 0 \le s \le t_m\}, \quad \textit{and} \quad \{W_s^c; t_m \le s \le T\}$$

are independent of each other, and the following equalities hold:

$$\begin{aligned} \{W_s^c\ ;\ 0 \le s \le t_m\} &=_d \sqrt{t_m}\left\{R_{\frac{t_m - s}{t_m}}(\delta_1)\ ;\ 0 \le s \le t_m\right\} + m \\ \{W_s^c\ ;\ t_m \le s \le T\} &=_d \sqrt{T - t_m}\left\{R_{\frac{s - t_m}{T - t_m}}(\delta_2)\ ;\ t_m \le s \le T\right\} + m, \end{aligned}$$

where $=_d$ *denotes equality in distribution,* $R(\delta)$ *is the Bessel Bridge between* 0 *and* $\delta \ge 0$ *of unit length, i.e.* $R_\delta = \{R_t(\delta);\ 0 \le t \le 1\}$*, and*

$$\delta_1 \doteq -\frac{m}{\sqrt{t_m}}; \qquad \delta_2 \doteq \frac{a - m}{\sqrt{T - t_m}}.$$

Proposition 2 allows to recover the path to be used as "proposals" in EA2, given the pair (m, t_m). All we need is random draws from the unit time length Bessel Bridge with the appropriate parameters. This can be easily done by using the following representation of the Bessel Bridge:

$$R_t(\delta) = \sqrt{(\delta t + BB_t^{(1)})^2 + (BB_t^{(2)})^2 + (BB_t^{(3)})^2},$$

where $BB^{(i)}$, $(i = 1, 2, 3)$ are independent standard Brownian Bridges, i.e., Brownian Bridge of unit time length starting and terminating at point 0. The representation formula reduces the problem of generating from the Bessel Bridge to the one of drawing from standard Brownian Bridge.

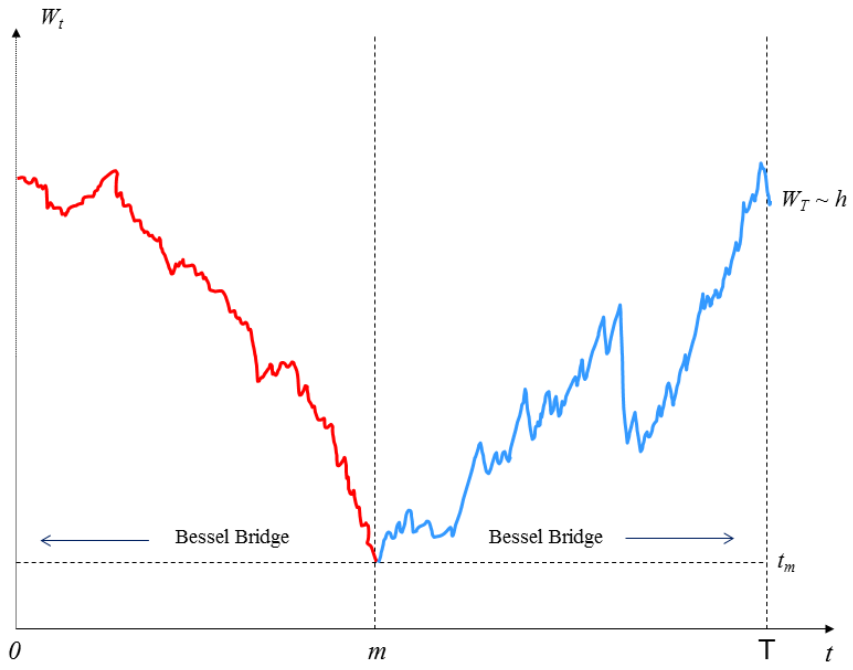

Figure 2.2: Path recostruction given minimum m and time it is achieved t_m.

The scheme described above provides a collection of instances (*skeleton*) of the target process at random times determined by the realization of the auxiliary Poisson process (plus the minimum of the path with the corresponding time). It is clear that, as in the case of EA1, realizations of the process at arbitrary times can be obtained from the dynamics of the bridge process used as proposal. Thus, in case of EA2, a realization of the process at time s, can be obtained by simulating from a distribution whose density can be expressed in terms of the transition density function $q(\cdot, \cdot, \cdot)$ of the Bessel process. In fact, in analogy with the EA1 case, let s_-, s_+ ($s_- < s < s_+$,), be the closest times on the left and on the right of s respectively, where the values of the sample path, say x_-, x_+, are already available. Then, using

proposition 2 and Markov property, it is easily shown that the density function of the random variable X_s (conditionally on the skeleton obtained from EA2) is:

$$\frac{q\left(\frac{s-s_-}{T-t_m},\ \frac{x_--m}{\sqrt{T-t_m}},\ \frac{x_s-m}{\sqrt{T-t_m}}\right)\times q\left(\frac{s_+-s}{T-t_m},\ \frac{x_s-m}{\sqrt{T-t_m}},\ \frac{x_+-m}{\sqrt{T-t_m}}\right)}{q\left(\frac{s_+-s_-}{T-t_m},\ \frac{x_--m}{\sqrt{T-t_m}},\ \frac{x_+-m}{\sqrt{T-t_m}}\right)\sqrt{T-t_m}}, \tag{2.28}$$

for $s > t_m$,

and:

$$\frac{q\left(\frac{s_+-s}{t_m},\ \frac{x_+-m}{\sqrt{t_m}},\ \frac{x_s-m}{\sqrt{t_m}}\right)\times q\left(\frac{s-s_-}{t_m},\ \frac{x_s-m}{\sqrt{t_m}},\ \frac{x_--m}{\sqrt{t_m}}\right)}{q\left(\frac{s_+-s_-}{t_m},\ \frac{x_+-m}{\sqrt{t_m}},\ \frac{x_--m}{\sqrt{t_m}}\right)\sqrt{t_m}}, \tag{2.29}$$

for $s \leq t_m$.

Funally, we recall that the exact algorithms described above allow to obtain realizations from the finite dimensional distributions of the transformed process $X = \eta(Y)$ where the function $\eta(\cdot)$ is given by formula (2.4). Thus, once realizations from the random vector $X_{t_1}, \ldots, X_{t_n}$ are available, realizations from $Y_{t_1}, \ldots, Y_{t_n}$ can be simply obtained by inverting the function η.

2.2.5 Efficiency of the exact algorithm

The efficiency of the exact algorithm can be studied in terms of the number of proposed paths that are generated until the first path is accepted. It is clear that the probability P of accepting a path depends on the number of points drawn from the auxiliary Poisson process, i.e., on the number of conditions to be checked. For the case of EA1, it follows from the definition of the Poisson process that the expected number D of the points on which the conditions for acceptance are to be checked is $M \times T$, where M is the upper bound for the functional Φ. From (2.15), using the Jensen inequality, it is easily shown that the acceptance probability is greater than $\exp(-M \cdot T)$. For EA2, it can be proved (Beskos, Papaspilioupolos and Roberts, 2006) that similar relations are valid with M replaced by $E[M(m_T)]$, where the r.v. m_T has been defined in Theorem 2.2.1. Moreover, it is not difficult to show

that, for both EA1 and EA2, the expectation of the total number of Poisson process points drawn until the first path is accepted is less than $E[D]/P$. The last result can be used to estimate the computational cost of the exact algorithm. For instance, for EA1, it implies that the computational cost is $O\left(MT \cdot e^{MT}\right)$. For EA2 it is difficult to find explicit bounds because one should find bound for $E[M(m_T)]$. However, it is clear that for both EA1 and EA2 in some cases it could be convenient to break the interval $[0, T]$ into (say K) smaller intervals in order to increase the efficiency. For instance, for EA1, this implies a computational cost $O\left(MT \cdot e^{MT/K}\right)$ that, for a fixed value of the ratio T/K, is linear in the length of the interval where the simulation is required.

2.2.6 Simulation of conditioned diffusions

The algorithms illustrated in the previous sections provide sample paths of a diffusion with a given starting point on a fixed interval $[0, T]$. In some situations, particularly in applications to inferential problems, one is interested in simulating sample paths conditioned on taking fixed values at both extremes of the time interval (diffusion bridges). It is important to stress that Eulerian schemes do not allow to obtain, even approximately, realizations from such conditioned diffusions. On the other hand it is well known (see, e.g., Rogers and D.Williams (2000)) that if Y is solution of the SDE (2.1), then the distribution of the diffusion Y conditioned on $Y_T = v$ is the same as that of a new process Y^v satisfying the SDE:

$$dY_t^v = \tilde{b}(t, Y_t^v)dt + \sigma(t, Y_t^v)dW_t, \qquad Y_0^v = y_0, \tag{2.30}$$

where:

$$\tilde{b}(t, y) \doteq b(t, y) + \sigma^2(t, y)\frac{\partial}{\partial y}\left\{\log p(t, y; T, v)\right\},$$

and $p(t, y; T, v)$ is the transition density of the (in general time non homogeneus) process Y between point (y) at time t and point v at time T.

Unfortunatly, formula (2.30) is not useful because in general (actually in all cases where simulation of the target process is not straightforward) the transition density p is not known.

Despite of this difficulties, a slight modification of the exact algorithm (actually a simplification) can be used to obtain exact simulations of diffusion bridges (of course for class of diffusion where EA is applicable). In fact,

reasoning as in Section 2.2.1, and decomposing the measures of the involved processes as products of the unconditional measures with the respective densities corresponding to the process values at the right exreme of the interval $[0, T]$, it can be argued that:

$$\frac{d\mathbb{P}^{t,x,y}}{d\mathbb{W}^{t,x,y}} = \exp\left\{A(y) - A(x) - \frac{1}{2}\int_0^T \left[\alpha(W_s)^2 + \alpha'(W_s)\right] ds\right\}, \quad (2.31)$$

where $\mathbb{P}^{t,x,y}$ and $\mathbb{W}^{t,x,y}$ denote the measures associated with the target bridge process and the Brownian Bridge respectively, both bridges being between points $(0, x)$ and (T, y). Formula (2.31) implies that the Radon Nikodym derivative of the measure $\mathbb{P}^{t,x,y}$ with respect to $\mathbb{W}^{t,x,y}$ is proportional to $\exp\left\{-\int_0^T \Phi(W_t)dt\right\}$, with Φ defined as in Section 2.2.2. The fact that in (2.31) the dominating measure is the one associated with the Brownian Bridge instead of the biased Brownian motion as in (2.14), allows to avoid the first step of algorithms EA1 and EA2 for simulation from the measure $\mathbb{P}^{t,x,y}$.

Remark

In case of conditional diffusion, it is sometimes convenient to describe the exact algorithm in an equivalent slightly different manner which will turn out to be useful when inference on diffusion parameters will be discussed. In fact, for both EA1 and EA2, we can define a (random, in general) quantity $r(W)$ such that:

$$r(W) \geq \sup_{s\in[0,T]} \frac{\alpha(W_s)^2 + \alpha'(W_s)}{2} - l.$$

Then, we simply re-define the functional Φ by scaling the definition (2.13):

$$\Phi(W_s) \doteq \frac{1}{r(W)}\left\{\frac{\alpha(W_s)^2 + \alpha'(W_s)}{2} - l\right\}, \quad (2.32)$$

so that, the Radon-Nikodym derivative (2.31) becomes:

$$\begin{aligned} \frac{d\mathbb{P}^{t,x,y}}{d\mathbb{W}^{t,x,y}} &= \exp\left\{A(y) - A(x) - r(W)\int_0^T \Phi(W_s)ds\right\} \quad (2.33) \\ &\propto \exp\left\{-r(W)\int_0^T \Phi(W_s)ds\right\} \leq 1. \end{aligned}$$

Note that, depending on $\alpha^2(\cdot)+\alpha'(\cdot)$ being boundend above or not (EA1 and EA2 cases respectively), $r(W)$ can be identified with the quantities M and $M(m)$ in paragraphs (2.2.3) and (2.2.4) respectively. From expression (2.33) it follows that in the EA algorithm, the Poisson process with unit intensity has to be replaced by a Poisson process with intensity $r(W)$. However, due to the re-definition of Φ, it is sufficient to draw from the Poisson process in the rectangular area $[0,T]\times[0,1]$ instead of $[0,T]\times[0,r(W)]$ as in steps 3 and 5 of EA1 and EA2 respectively.
As final remark, we note that in cases of diffusion bridges, although the exact algorithm is simpler than the corresponding algorithm for unconditioned diffusions, the computational cost cannot be reduced by breaking the interval $[0,T]$ in sub-intervals. This poses some practical restrictions to the applicability of EA in inferential problems.

2.2.7 Extensions

In this Section, possible extensions of the exact algorithm are presented. The first extension discussed concerns time non-homogeneous diffusions. Let us consider the SDE (2.1), where drift and diffusion $b\equiv b(t,y)$ and $\sigma\equiv\sigma(t,y)$ are allowed to explicitly depend on time. Note that also in this case we can reduce the initial SDE to one with unit volatility through the transformation:

$$X_t=\eta(t,Y_t),\quad \text{where}\quad \eta(t,y)=\int_c^y\frac{1}{\sigma(t,u)}du,$$

where in this setting the function η explicitely depends on time.
Starting from the Girsanov theorem and using Ito's rule, it can be easily proved that an acceptance-rejection scheme can be defined in terms of a new functional Φ which is the generalization of the one defined in Section 2.2.2:

$$\begin{aligned}\Phi(t,x) &\doteq \frac{1}{2}\left\{\alpha^2(t,x)+\alpha'(t,x)+2\frac{\partial A(t,x)}{\partial t}\right\}-l,\\ l &\doteq \inf_{t\in[0,T]}\inf_{x\in\mathbb{R}}\left\{\frac{1}{2}\alpha^2(t,x)+\frac{1}{2}\alpha'(t,x)+\frac{\partial A(t,x)}{\partial t}\right\},\end{aligned}$$

where $A(t,x)\doteq\int_c^x\alpha(t,u)du$, and $\alpha(t,x)$ is the drift of the SDE satisfied by the transformed process X:

$$\alpha(t,x) = \frac{b(t,\eta^{-1}(t,x))}{\sigma(t,\eta^{-1}(t,x))} - \frac{1}{2}\frac{\partial\sigma}{\partial y}(t,\eta^{-1}(t,x)).$$

A more involved issue is the extension of the methodology to multivariate contexts. In fact, let $Y : \Omega \times [0,T] \to \mathbb{R}^d$ be a strong solution of the SDE

$$dY_t = b(Y_t)dt + \Sigma(Y_t)dW_t \qquad Y_0 = y_0, \tag{2.34}$$

where $b(\cdot) : \mathbb{R}^d \to \mathbb{R}^d$ is a d dimensional drift, Σ is a $d \times d$ matrix valued function on $\mathbb{R}^d$, and W is a d-dimensional Brownian motion. The first problem we encounter is to find a function $\eta : \mathbb{R}^d \to \mathbb{R}^d$ such that the volatility of the process X defined via the transformation $X_t = \eta(Y_t)$ is the identity matrix in $\mathbb{R}^d$. From the multidimensional Ito's lemma it follows that the function $\eta(\cdot) \equiv \left(\eta^1(\cdot), \ldots, \eta^d(\cdot)\right)$ must fulfill the conditions:

$$\nabla\eta^j = \Sigma^{-1}_{j+} \qquad j = 1, \ldots, d \tag{2.35}$$

where Σ^{-1}_{j+} is the jth row of the matrix Σ^{-1}. In other words, the rows of the matrix Σ^{-1} are vector fields of gradient type (conservative fields). If the d conditions (2.35) are satisfied, then one can define the new d-dimensional process $X \equiv (X^1, \ldots, X^d)$ as:

$$X^j_t = \eta^j(Y_t), \qquad \text{where } \eta^j(y) \doteq \int_c^y \Sigma^{-1}_{j+}(u) \cdot du \quad (j = 1, \ldots, d),$$

the integral being independent of the particular path in $\mathbb{R}^d$ from the (arbitrary) point c to the point y. From standard conditions for a vector field to be conservative, it follows that trasformation of the original SDE (2.34) in a new SDE with identity diffusion matrix is possible if and only if:

$$\frac{\partial\Sigma_{ij}(y)}{\partial y_k} = \frac{\partial\Sigma_{ik}(y)}{\partial y_j}, \quad \forall i,j,k = 1, \ldots, d. \tag{2.36}$$

Matrices Σ that verify the conditions (2.36) are sometimes referred to as *reducible* (Ait-Sahalia, 2008).

A second problem we have to deal with in extending the methodology to multivariate diffusions concerns the possibility of eliminating the stochastic integral in the expression of the Radon-Nykodyn derivative of the target measure with respect to the measure associated with the biased Brownian motion (or with the Brownian Bridge in the conditional case). In fact, once the initial SDE has been reduced to the standardized form (provided it is possible)

$$dX_t = \alpha(X_t)dt + dW_t \qquad X_0 = x_0, \tag{2.37}$$

generalization of formulas (2.8)-(2.10) implies that the d-vector valued function $\alpha(\cdot)$ must be of gradient type, i.e., $\alpha(x) = \nabla A(x)$ for some scalar field A on $\mathbb{R}^d$. In this case, the integral (2.9) is to be intended as a curvilinear (path independent) integral.

Chapter 3

Parametric inference for discretely observed diffusion processes

3.1 Introduction

The last three decades have seen a tremendous explosion of research in the area of statistical inference for diffusion processes. Interest in this topic arises from the fact that diffusions have been increasingly used for modelling a big variety of phenomena randomly varying in time. In fact, they are flexible and powerful tools for the analysis of phenomena whose evolution can be tought of as continuous in time. From a practical point of view, continuous time is a useful assumption whenever processes are observed at high frequency or, in any case, when lag between pairs of consecutive observations is not a known fixed time interval. For this feature, diffusions have been widely used in mathematical finance to describe dynamics of stock prices, interest rates, exchange rates, etc. Although these quantities are modeled as continuously varying in time, they are typically measured at discrete times. For instance, discrete observations may correspond to months, weeks, days, or even to random times. Thus, the problem arises of making inference on the generating process based on a set of discrete observations. The problem of estimation for diffusion processes observed at discrete times is highly non-trivial. This is because diffusions are often specified as solutions of stochastic differential equations (SDEs), and the transition density is typically intractable. Important exceptions where the transition density can be explicitly known are solutions of linear SDEs (e.g., Brownian motion with drift and Ornstein Uhlenbeck process) having Gaussian transition density, or some other special

cases, such as the geometric Brownian motion or the Cox-Ingersol-Ross (CIR) model, whose transition densities are log-normal and non-central chi-square respectively. These simple models have been extensively used in many areas of reasearch, but they are sometimes not adequate to describe the observed data. For instance, it has been recognized that log-prices do not seem to be independent and Gaussian as assumed in market models based on the geometric Brownian motion (Black and Schole theory). Nevertheless, as we will see in the following, transition densities of diffusions of interest can often be expressed in terms of the known transition densities of these simple processes. From an inferential point of view, non-tractability of transition density determines difficulties for approaches based on likelihood function. This has motivated significant research focused on methods for likelihood approximation (both analytical and numerical), or alternative approaches. In this chapter we will give a short overview of the main inferential approaches limiting ourselves to parametric inference and mainly focusing on methods based on numerical approximations of the transition densities for one-dimensional processes.

3.2 Model and set-up

In the following, and throughout this chapter, we will consider the canonical set-up in which a model on the filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ is specified through the one-dimensional time homogeneous SDE:

$$dY_t = b(Y_t, \theta)dt + \sigma(Y_t, \theta)dW_t, \quad Y(0) = y_0, \tag{3.1}$$

where W is a standard Brownian motion and θ is a p-dimensional parameter in some subset $\Theta \subset \mathbb{R}^p$. The drift and the diffusion coefficients $b, \sigma : (\mathbb{R} \times \Theta) \to \mathbb{R}$ are assumed to be smooth functions of their arguments and the true value of the parameter is denoted by θ_0. Moreover, it is assumed that equation (3.1) has a unique strong solution Y and a stationary distribution $f_{\theta_0} : \mathbb{R} \to \mathbb{R}$ exists, i.e., the process Y is ergodic. These assumptions are quite standard and are necessary to ensure asymptotic consistency properties of various estimators.
An important issue in the study of parametric inference on diffusion processes is the sampling scheme, i.e., the way the observations are available. Different asymptotic contexts are analyzed in the literature depending on the approach followed. In particular, if $t_1, \ldots, t_n$ are the times when observations are available, the limit $n \to \infty$ can be analized in the set-up where

$\max_{i=1,\ldots,n} |t_{i+1} - t_i| \to 0$ and $t_n = T$ (T fixed), or from the different perspective where the time points are equally spaced, i.e., $t_i = t_0 + i\Delta$, (Δ fixed). One reason of interest for the sampling scheme arises from the fact that for many estimation approaches, estimators of parameters involved in the drift and diffusion coefficients have quite different properties. In fact, if a sample path were entirely known on the interval $[0, T]$ (continuous observations), then, in principle, the diffusion cofficient could be *exactly* known as well, being related to the quadratic variation of the process (Karatzas and Shreve, 1991). More precisely, given the partition $\Pi \equiv (t_0, t_1, \ldots, t_n)$ of $[0, t]$, with $t_0 \leq t_1 \leq \ldots, \leq t_n = t$, and defining the p-th variation over Π of the process Y as:

$$V_t^{(p)}(\Pi) = \sum_{1=1}^{n} |Y_{t_i} - Y_{t_{i-1}}|^p, \tag{3.2}$$

we have:

$$\lim_{\|\Pi\| \to 0} V_t^{(2)}(\Pi) = \langle Y \rangle_t \quad \text{in probability}, \tag{3.3}$$

where $\|\Pi\|$ is the *mesh* of the partition Π, i.e., $\|\Pi\| = \max_i |Y_{t_i} - Y_{t_{i-1}}|$, and $\langle Y \rangle_t$ denotes the quadratic variation of Y:

$$\langle Y \rangle_t = \int_0^t \sigma^2(Y_s, \theta) ds.$$

Thus, intuitively, sampling at higher and higher frequency allows an increasingly accurate estimate of σ. In contrast, estimation of drift does not have this property, and tipically consistency requires availability of data on increasingly longer time intervals.

3.3 Continuous time likelihood

In this Section, as usually, $(\mathcal{C}[0, T], \mathcal{A})$ will denote the space of the continuous functions on $[0, T]$ equipped with the σ-algebra of the cylinder sets, and ω a

typical element of $\mathcal{C}[0,T]$. We refer to the usual framework where the target diffusion process Y induces a probability measure on $\mathcal{C}[0,T]$, such that the coordinate mapping process defined via $Y_t(\omega) = \omega(t)$ has the law of Y. In this framework, as we have already mentioned in Chapter 2, an important consequence of the Girsanov theorem on the change of meausure is that, under certain conditions, measures associated to solutions of SDEs with the same diffusion term are absolutely continuous with respect to one another (Oksendal, 1998). More in detail, let Y be a solution of the SDE

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t,$$

and X a solution of another SDE having the same diffusion function $\sigma(\cdot)$ but different drift, say $a(\cdot)$ (here dependence on parameter has been removed from the notation), and let $\mathbb{P}_Y$, $\mathbb{P}_X$ be the measures on $(\mathcal{C}[0,T], \mathcal{A})$ associated with processes Y and X respectively. Then, the Girsanov theorem implies that the Radon-Nikodyn derivative of the measure $\mathbb{P}_Y$ w.r.t. $\mathbb{P}_X$ is:

$$\frac{d\mathbb{P}_Y}{d\mathbb{P}_X} = \exp\left\{\int_0^t h(Y_s)dW_s - \frac{1}{2}\int_0^t h^2(Y_s)ds\right\}, \tag{3.4}$$

where function h satisfies the equation:

$$\sigma(y)h(y) = b(y) - a(y).$$

For inferential purposes, the Radon-Nykodin derivative (3.4) is more conveniently written as:

$$\exp\left\{\int_0^t \frac{b-a}{\sigma^2}(Y_s)dY_s - \frac{1}{2}\int_0^t \frac{b^2-a^2}{\sigma^2}(Y_s)ds.\right\}, \tag{3.5}$$

Now, assume that the diffusion coefficient is known and interest is in estimating parameters in the drift term $b \equiv b(y,\theta)$. Then, using (3.5) with $a \equiv 0$ provides an expression $L_t^c(\theta)$ that can be interpreted as *continuous time likelihood* for the parameter θ (Lipster and Shiryayev, 1977) :

$$L_t^c(\theta) = \exp\left\{\int_0^t \frac{b(Y_s,\theta)}{\sigma^2(Y_s)}dY_s - \frac{1}{2}\int_0^t \frac{b^2(Y_s,\theta)}{\sigma^2(Y_s)}ds.\right\}, \tag{3.6}$$

where dependence on θ is only in the drift coefficient. When data $Y_{t_1}, \ldots, Y_{t_n}$ are available at discrete time points $t_1, \ldots, t_n = t$, and when σ is assumed to be known, a possible approach is to consider the log-likelihood obtained from Ito-Riemann approximation of (3.6):

$$L_t^{disc}(\theta) = \sum_{i=1}^{n} \left\{ \frac{b(Y_{t_i}, \theta)}{\sigma^2(Y_{t_i})}(Y_{t_{i+1}} - Y_{t_i}) - \frac{1}{2}\frac{b^2(Y_{t_i}, \theta)}{\sigma^2(Y_{t_i})}\Delta t_i \right\}, \tag{3.7}$$

where $\Delta t_i = t_{i+1} - t_i$.

Note that the approximate likelihood (3.7) is, up to a multiplicative constant, the *exact* discrete-time likelihood of the approximate model obtained via Euler-Maruyama scheme:

$$\Delta Y_i \doteq Y_{t_{i+1}} - Y_{t_i} = b(Y_{t_i}, \theta)\Delta t_i + \sigma(Y_{t_i}, \theta)\sqrt{\Delta t_i} Z_i, \tag{3.8}$$

where $Z_i \sim N(0, 1)$ are independent standard Gaussian variables. Model (3.8) has been studied by Florens-Zmirou (1989) in the case of constant diffusion term, $\sigma(Y_t) \equiv \sigma$. In this case, Prakasa-Rao (1983) studied the (least squares) estimator of θ obtained by minimizing (3.7). The previous approach has been extended by Yoshida (1992) and Florens-Zmirou (1989) to the case of unknown but multiplicative diffusion coefficient: $\sigma(x, \psi) = \psi g(x)$ where ψ is a parameter appearing only in σ and g is a real valued function. In their works, ψ is estimated using the relation between σ and the quadratic variation, and θ is still estimated by minimizing (3.7). Florens-Zmirou (1989) proves asymptotic efficiency in the context of fixed time intervals $h_n \doteq t_{i+1} - t_i$, $(i = 1, \ldots, n)$ between consecutive data points under the assumption $nh_n^2 \to 0$ (*rapidly increasing experimental design*). Using a Gaussian approximation of the transition density more accurate than the one implied by (3.7), Kessler (1997) obtaines joint estimates of the parameters in b and σ by assuming $nh_n^p \to 0$, with p arbitrary integer.

One problem with the discrete approximation of the continuous time likelihood, or equivalently with the approximation of the target process via Euler-Maruyama scheme, is that the resulting estimators are strongly biased for not sufficiently small intervals between consecutive observations. In Florens-Zmirou (1989) it is shown that, for fixed $\Delta = t_{i+1} - t_i$, $(i = 1, \ldots, n)$, the estimator is not consistent for $n \to \infty$.

3.4 Discrete time likelihood

From now on we will consider time homogeneous diffusion processes specified through the SDE (3.1), with transition density $p(t, x, y; \theta)$ defined via:

$$P\{Y_{s+t} \in \Gamma | Y_s = x; \theta\} = \int_\Gamma p(t, x, y; \theta) dy, \quad \forall s > 0, \quad \Gamma \in \mathcal{F}_t$$
$$p(0, x, y; \theta) = \delta(y - x).$$

We assume that the process Y is observed at a finite collection of times $0 = t_0, t_1, \ldots, t_n = t$, the corresponding values being $\mathbf{y} = (Y_0 = Y_{t_0}, Y_{t_1}, \ldots Y_{t_n} = Y_t)$, i.e., Y is discretely observed. From Markov property it follows that the log-likelihood based on the n observations Y_{t_i} can be written as:

$$l(\theta|\mathbf{y}) = \sum_{i=1}^{n} \log p(\Delta_{t_i}, Y_{t_i}, Y_{t_{i+1}}; \theta), \tag{3.9}$$

where $\Delta_{t_i} = t_{i+1} - t_i$.
Under mild conditions, the estimator obtained by maximizing the function (3.9) has the usual nice consistency properties, but unfortunately, except for a few cases, the transition density $p(t, x, y)$ is unknown. Among processes for which transition density is explicitly known, an important class is composed of solutions of linear SDEs, i.e., SDEs of the form:

$$dY_t = (b(t)Y_t + a(t))\, dt + \sigma(t) dW_t, \quad Y(0) = y_0, \tag{3.10}$$

with a, b, σ real valued functions. Strong solutions of (3.10) are known to be Gaussian processes with mean m_t and variance V_t which satisfy the ordinary differential equations (ODE):

$$\begin{aligned} \frac{dm_t}{dt} &= b(t)m_t + a(t), & m_0 = y_0, \\ \frac{dV_t}{dt} &= 2b(t)V_t + \sigma^2(t), & V_0 = 0. \end{aligned}$$

In the case b, a, σ are numerical constants, the solution of (3.10) is the well known Ornstein-Uhlenbeck process, whose transition density is:

$$p(t,x,y;\theta) = \frac{1}{\sqrt{2\pi V_t}} \exp\left\{ \left[y - x - \frac{a}{b}\left(e^{bt}-1\right)\right]^2 / (2V_t) \right\}$$

where:

$$V_t = \frac{\sigma^2}{2b}\left(e^{bt}-1\right),$$

and $\theta \equiv (b, a, \sigma)$.

3.4.1 Analytic approximations and estimating functions

Based on log-likelihood (3.9) with the above transition density function, maximum likelihood estimates (MLEs) can be easily found. Because of the easiness in using discrete likelihood approach for the linear case, many methods for inference on diffusions are based on linearization of the corresponding SDEs. One example is the Euler-Maruyama scheme described in the previous Section. This scheme results in approximating conditional distributions with Gaussian distributions and can determine serious bias if time intervals between consecutive observations are not small. Using a more sophisticated approximation of the transition density based on Hermite polynomial expansion, Ait-Sahalia (2002) proposes an estimator which converges to the maximum likelihood estimator as the expansion order increases. A possible alternative approach is based on *estimating functions* (Sorensen, 1999*a*,*b*; Heyde, 1997). In the context where the parameter θ to be estimated is p-dimensional, an estimating function is a $\mathbb{R}^p$-valued function $F(y_1, \ldots, y_n; \theta)$ with data and parameter as its arguments. An estimator $\hat{\theta}$ of θ is obtained by solving the equation $F(y_1, \ldots, y_n; \theta) = 0$ w.r.t θ. The reference example of estimating function is the score function. In problems where it is not available, as in the case of diffusions with unknown transition density, one tries to find an estimating function which is able to distinguish the "true value" of the parameter θ_0 from all other values:

$$E_{\theta_0}\left[F_n(y_1, \ldots, y_n; \theta)\right] = 0 \quad \text{iff} \quad \theta = \theta_0. \tag{3.11}$$

A common choice is to take an estimating function which is a martingale with respect to the discrete-time filtration generated by the data. In case of fixed time interval Δ, the estimating function $F_n(\theta)$ (we have removed dependence on data from the notations) can be written as:

$$F_n(\theta) = \sum_{i=1}^{n} f(Y_{(i-1)\Delta}, Y_{i\Delta}, \theta),$$

where f verifies : $E_{\theta_0} f(Y_0, Y_\Delta, 0) = 0 \quad$ if and only if $\quad \theta = \theta_0$. The advantage of using martingales as estimating functions is that the properties of the resulting estimators can be derived from the classical asymptotic theory for stationary martingales (Billingsley, 1961). Another common choice is to use *simple functions* (Kessler, 2000), i.e., functions of the form:

$$F_n(\theta) = \sum_{i=1}^{n} f(Y_{i\Delta}, \theta).$$

In this case, the condition (3.11) for distinction of the true parameter θ_0 simplifies to: $E_{\theta_0} f(Y_0, \theta) = 0 \quad$ if and only if $\quad \theta = \theta_0$. Here we assume that $Y_0 \sim \mu_{\theta_0}$, μ_{θ_0} being the invariant stationary distribution of the (ergodic) process Y.

3.4.2 Numerical solutions of Fokker-Planck equation

It is well known that, under some regularity conditions, for fixed x, θ the transition density $p(t, x, y; \theta)$ of the SDE (3.1) is solution of the partial differential equation (PDE) called Fokker-Planck equation, or Kolmogorov forward equation (Karatzas and Shreve, 1991):

$$\frac{\partial}{\partial t} p(t, x, y; \theta) = -\frac{\partial}{\partial y}(b(y; \theta) p(t, x, y; \theta)) + \frac{1}{2} \frac{\partial^2}{\partial y^2}(\sigma^2(y; \theta) p(t, x, y; \theta)), \quad (3.12)$$

with initial condition $p(0, x, y; \theta) = \delta(y - x)$. Actually, n equations of type (3.12) have to be solved, one for each value of $x = Y_{t_i}$, and what is of interest for the likelihood function is the set of values that the function $(t, y) \rightarrow p(t, Y_{t_{i-1}}, y; \theta)$ takes on the points (Δ_{t_i}, Y_{t_i}). Typically, equation (3.12) must be solved numerically through some finite difference scheme.

It can be proved that, maximization of the approximate log-likelihood obtained from a suitable discretization of equation (3.12), results in an estimator asymptotically equivalent to the true MLE, provided that the step of the lattice used for discretization tends to zero sufficiently fast (Poulsen, 1999).

3.4.3 Estimators based on indirect inference

Another approach to parametric inference on diffusion processes is based on using *indirect inference* (Gourieroux et al., 1993). According to this approach, an auxiliary model, easy to estimate, is fitted to the data $Y_{t_1}, \ldots, Y_{t_n}$. Assume that the auxiliary model is characterized by a (possibly multi dimensional) parameter ρ and that maximum likelihood estimate $\hat{\rho}_n$ based on the n sample data is available. Then, for different values of parameter θ, a long trajectory $Y'_{t_1}, \ldots, Y'_{t_m}$ is simulated from the original model and, based on each trajectory corresponding to a specific value of θ, an MLE estimator $\hat{\rho}_m(\theta)$ is calculated. Finally, an estimator of θ is obtained by minimizing the quadratic form:

$$[\hat{\rho}_n - \hat{\rho}_m(\theta)]^t \, \Omega \, [\hat{\rho}_n - \hat{\rho}_m(\theta)] \,,$$

where Ω is a suitably defined semi-positive definite matrix.

A natural choice is to use, as auxiliary model, the one obtained by approximating the solution of the generating SDE via Euler-Maruyama scheme. This results in estimating parameters of Gaussian distributions and then adjusting for the discrepancies deriving from the assumption of normality of the process increments implied by the Eulerian approximation.

3.5 MC approximation of the transition density function

3.5.1 Introduction

An important class of inference methods for diffusions observed at discrete times is based on the numerical approximation of their transition density

via Monte Carlo simulation. Within this class, two broad strategies can be distinguished. The first one is based on approximating the process itself with a finite dimensional random variable via some Euler-type discretization scheme and estimating the transition density corresponding to each pair of time-consecutive observations via a large number of MC simulations of the approximate process. This approach is sometimes referred to as a *projection-simulation* approach (Papaspiliopoulos and Roberts, 2009). The problem with this approach is that the number n of intermediate points between consecutive observations used in discretization has to be taken large enough to reduce bias, typically implying increase in variance.
The second strategy (*simulation-projection*) more relies on the infinite-dimensional nature of the problem and tries to directly obtain (at least in principle) realizations from the (infinite-dimensional) distribution of the process conditional on pairs of observed values (diffusion bridges). This approach is mathematically more complex in that it involves change of measure in infinite-dimensional spaces. Again, simulated diffusion bridges have to be approximated by n-tuples of intermediate points and the resulting bias tends to zero as $n \to \infty$. However, trade-off between bias and variance is typically better than with projection-simulation strategy, and for some class of models, the fine discretization step can be avoided using the restrospective-sampling approach (see Chapter 1). The simulation-projection approach can be viewed as a strategy for a missing data problem where missing data can be identified with the unobserved path between (discrete) observations. For this reason, different methods in this framework are related, to some extent, to the problem of simulating diffusion bridges (*imputation*). This problem is difficult to treat in general, and in particular in the framework of Eulerian-type schemes, which are typically not able to take into account conditioning on ending point.

A more appropriate approach seems to be the one based on change of measure in the space of sample paths. This approach naturally leads to acceptance-rejection and importance sampling schemes that can be incorporated in the inferential procedure. In the remainder of this Section we will provide some examples of inferential procedures based on both the approaches, with major emphasis on the simulation-projection approach. In the following we refer to the context of Section 3.4 and assume that the log-likelihood (3.9) is to be estimated. In the homogeneous case we can limit ourselves to the analysis of a single term in the sum (3.9). Thus, we assume that the target quantity to estimate is the transition density function $p(t, x, y; \theta)$, from point $(0, x)$, to (t, y).

3.5.2 Approach based on simulation of unconditional paths

A first proposal for inference based on simulation of diffusion paths unconditionally on the ending point is due to Pedersen (1995). The main idea is to approximate the target diffusion on the interval $[0, t]$ through an Eulerian scheme based on m points corresponding to intermediate times in $[0, t]$, the first point being x at time 0. Let Y^m be the corresponding approximate process, and assume for simplicity that the time-intervals between consecutive points have the same length $\Delta_m = t/m$. For a single realized trajectory of Y^m, let y_{m-1} be the point corresponding to time $t(m-1)/m$ (the last point before y). Then, if $\tilde{p}_m(s, u, v; \theta)$ denotes the transition density of the approximate process Y^m, we have, from the Chapman-Kolmogorov formula:

$$\begin{aligned} \tilde{p}_m(t, x, y; \theta) &= \int \tilde{p}_m((m-1)\Delta_m, x, y_{m-1}; \theta)\tilde{p}_m(\Delta_m, y_{m-1}, y; \theta)dy_{m-1} \\ &= E_\theta^m \left[\tilde{p}_m((m-1)\Delta_m, x, y_{m-1}; \theta)|Y_0 = x\right], \end{aligned} \tag{3.13}$$

where E_θ^m denotes expectation w.r.t. the law of the approximate process corresponding to parameter θ. Now, for large m, the last term of (3.13) should be close to the analogue expression obtained by replacing E_θ^m with the expectation E_θ w.r.t. the law of the original process Y, and $\tilde{p}_m((m-1)\Delta_m, x, y_{m-1}; \theta)$ with the target transition density corresponding to the same points x and y_{m-1}. This suggests that one could estimate $p(t, x, y; \theta)$ through the Monte Carlo average, over a large number N of simulations, of the quantities:

$$\tilde{p}_m(\Delta_n, y_{m-1}^k, y; \theta) = \frac{1}{\sqrt{2\pi\Delta \cdot \sigma^2(y_{m-1}^k, \theta)}} \exp\left\{\frac{(y - y_{m-1}^k - \Delta \cdot b(y_{m-1}^k))^2}{2\Delta \cdot \sigma^2(y_{m-1}^k, \theta)}\right\},$$

where, for $k = 1, \ldots, N$, y_{m-1}^k is the k-th realization from the marginal distribution of the approximate process Y^m corresponding to time $(m-1)\Delta$.

It can be proved (Pedersen, 1995) that, considering each pair of consecutive observed points $(Y_{t_{i-1}}, Y_{t_i})$ as points x, y above (i.e., $x = Y_{t_{i-1}},\ y = Y_{t_i}\ t = t_i - t_{i-1} = \Delta_{t_i}$), the approximate likelihood:

$$L_n^m(\theta) = \prod_{i=1}^{n} \tilde{p}_m(\Delta_{t_i}, Y_{t_{i-1}}, Y_{t_i}; \theta), \tag{3.14}$$

converges in probability to the true likelihood:

$$L_n(\theta) = \prod_{i=1}^{n} \tilde{p}(\Delta_{t_i}, Y_{t_{i-1}}, Y_{t_i}; \theta),$$

and that there exists a sequence $m(n)$ such that, maximizing $L_n^{m(n)}(\theta)$ provides an estimator of θ which is asymptotically equivalent to the MLE. However, MC approximation of approximate likelihood (3.14) typically requires a large number of simulations, because of the lack of efficiency deriving from ignoring the realized values of the process on the right extreme of the intervals, in the path simulation.
Essentially, the same approach based on simulating paths unconditionally on the ending points is followed by Beskos, Papaspilioupolos and Roberts (2006), but in this case trajectories are drawn from the *exact* probability distribution associated with the driving SDE using the retrospective sampling technique (see Chapter 1). Being based on exact simulation, the methodology does not require drawing instances of the processes on a very fine time-grid, but its applicability relies on the conditions under which the EA is feasible. In detail, suppose that, for a fixed value of the parameter θ, we want to estimate the transition density function $p(t, x, y; \theta)$, where t is the time-separation between the two consecutive observations x and y of the target process Y. Due to time-homogeneity, we can take, without loss of generality, : $x = Y_0$, $y = Y_t$. The first step (ee Chapter 2) is to "standardize" the process through the transformation:

$$X_s = \eta(Y_s; \theta); \qquad \eta(y; \theta) = \int^{y} \frac{1}{\sigma(u, \theta)} du, \tag{3.15}$$

so that the new process satisfies the following SDE with unit diffusion coefficient:

$$dX_t = \alpha(X_t, \theta)dt + dW_t, \qquad X_0 = \eta(x; \theta), \tag{3.16}$$

where

$$\alpha(u, \theta) = \frac{b(\eta^{-1}(u; \theta), \theta)}{\sigma(\eta^{-1}(u; \theta), \theta)} - \frac{1}{2}\sigma'(\eta^{-1}(u; \theta), \theta) \tag{3.17}$$

A change of measure argument shows that the target transition density p is related to the transition density $\tilde{p}$ associated with the process X, via the relation:

$$p(t, x, y; \theta) = \tilde{p}(t, \eta(x; \theta), \eta(y; \theta); \theta)|\eta'(y; \theta)| \tag{3.18}$$

Thus, we are now interested in estimating the function $\tilde{p}(t, \eta(x; \theta), \eta(y; \theta); \theta)$ for a specific value of θ. To this aim, we can apply the EA (in the version EA1 or EA2 depending on the boundness conditions of the drift of transformed process, see Chapter 2) in order to obtain a skeleton of the process X in a time interval $[0, t+\delta]$ that includes the time t at which the value y is observed for the original process. The skeleton is composed of a collection of instances of the process $(X_0, X_{\tau_1}, \ldots, X_{t+\delta})$ corresponding to times $(0, \tau_1, \tau_2, \ldots, , t+\delta)$, where $(\tau_1, \tau_2, \ldots,)$ are random. Let ω be a path between the times $s = 0$ and $s = t + \delta$ accepted by the acceptance-rejection scheme used in the EA, and $S(\omega)$ its skeleton provided by the algorithm. It is clear that $S(\omega)$ is non-empty because it contains at least the pairs $(0, X_0)$ and $(t + \delta, X_{t+\delta})$. Then, conditionally on the current skeleton, the transition density $\tilde{p}(t, X_0, X_t|S)$ of the random variable X_t (conditional one-dimensional distribution of the process X at time t) is known, being related to the Brownian Bridge or Bessel Bridge dynamics depending on the particular algorithm (EA1 or EA2) being used. In particular, as illustrated in Chapter 1, due to the Markov property, it is sufficient to consider conditioning on the points of the skeleton (t_-, X_{t_-}), (t_+, X_{t_+}), where t_- and t_+ are the closest times to t, on the left and on the right respectively, returned by the EA. Thus, for instance, in the case of EA1 we have:

$$\tilde{p}(t, \eta(x; \theta), \eta(y; \theta)|S(\omega)) = \frac{1}{\sqrt{2\pi\Sigma}} exp\left\{\frac{1}{2\Sigma}(\eta(y; \theta) - \mu)^2\right\},$$

where:

$$\mu \equiv \mu(X_{t_-}, X_{t_+}, t_-, t_+) = X_{t_-} + \frac{t - t_-}{t_+ - t_-}(X_{t_+} - X_{t_-})$$

and:

$$\Sigma \equiv \Sigma(X_{t_-}, X_{t_+}, t_-, t_+) = \frac{(t - t_-)(t_+ - t)}{(t_+ - t_-)}.$$

Note that, in case of EA2, the skeleton includes also the minimum of the path and the time at which it is attained. The target transition density can be expressed as:

$$\tilde{p}(t, \eta(x;\theta), \eta(y;\theta);\theta) = E_S\left[\tilde{p}(t, \eta(x;\theta), \eta(y;\theta);\theta)\ |S\right]. \tag{3.19}$$

where E_S denotes expectation with respect to the law of S. The latter is in general intractable, but the r.h.s. of (3.19) can be approximated via MC simulation. This method, called *bridge method* in Beskos, Papaspiliopoulous and Roberts (2006) is based on exact simulation, thus the estimate of the transition density is unbiased for each value of θ and the only approximation is in MC average of conditional transition densities. However, its applicability relies on the feasibility of the EA. Moreover, as for the methodology in Pedersen (1995), drawing paths unconditionally on the observed value on the right extreme of the interval can result in inefficient estimation.

3.5.3 Methods based on diffusion bridge simulation

Lack of efficiency due to unconditional simulation in the bridge method can be avoided by using inferential methods which take into account all the information in observed data, i.e., the values of the process on both the extremes of the time interval between consecutive observations. As a matter of fact, these methods involve to some extent simulation of diffusion bridges. The core of the methodology is based on the change of measure for diffusion bridges. From now on, we can consider diffusions which are solutions of SDEs of the form (3.16), i.e., with unit volatility term. In fact, the process Y solution of the SDE (3.1) can always be transformed in the process X solution of (3.16), via the tranformation (3.15). Moreover, by means of the change of variable formula (3.18), the problem of estimating $p(t, u, v;\theta)$, with $Y_0 = u,\ Y_t = v$, is the same as that of estimating $\tilde{p}(t, x, y;\theta)$, where $x = X(0) = \eta(u,\theta)$, and $y = X(t) = \eta(v,\theta)$. Thus, we focus on estimation of $\tilde{p}(t, x, y;\theta)$, keeping in mind, however, that the values x and y are not directly observed, but depend on the observed data u and v, and on the parameter θ through the transformation (3.15). In the following, dependence on θ of initial and final point will not be reported. The estimation of the transition density of the tranformed process X is again based on the change of measure on the space $(\mathcal{C}[0,t], \mathcal{A})$. An intuitive argument is based on the Girsanov theorem and on the factorization of the measure associated with the diffusion process $\{X_s, 0 \le s \le t\}$ as product of a bridge measure and the measure on the real

line corresponding to the law of the random variable X_t. Specifically, let $\mathbb{P}_\theta^x$ be the measure associated with process X (with initial point x) on $[0,t]$, and $\mathbb{P}_\theta^{t,x,y}$ the measure associated with X conditioned on having its ending point y at time t. Then, we have the factorization: $\mathbb{P}_\theta^x = \mathbb{P}_\theta^{t,x,y} \otimes \tilde{p}(t,x,y;\theta) \cdot Leb(y)$, where $Leb(y)$ denotes the Lebesgue measure on $\mathbb{R}$. A similar decomposition holds for the Wiener measure $\mathbb{W}^x$ corresponding to the Brownian motion starting from x: $\mathbb{W}^x = \mathbb{W}^{t,x,y} \otimes \mathcal{N}(y-x,t) \cdot Leb(y)$, where $\mathbb{W}^{t,x,y}$ is the the measure induced by the appropriate Brownian bridge, and $\mathcal{N}(y-x;t)$ is the density of a zero-mean Gaussian r.v. with variance t. From the Girsanov theorem, we have:

$$\frac{d\mathbb{P}_\theta^x}{d\mathbb{W}^x} = Z_t(x),$$

where

$$Z_t^x(W) = \exp\left\{ \int_0^t \alpha(W_s^x;\theta) dW_s^x - \frac{1}{2}\alpha^2(W_s^x;\theta) ds \right\}.,$$

It follows:

$$\mathcal{N}(y-x,t) Z_t^x(W) = \tilde{p}(t,x,y;\theta) \frac{d\mathbb{P}_\theta^{t,x,y}}{d\mathbb{W}^{t,x,y}}(W) \tag{3.20}$$

Letting

$$A(x,\theta) = \int_c^x \alpha(u;\theta) du,$$

with c arbitrary constant, and using the Ito's formula as in formula 2.10 of Chapter 2 in order to eliminate the stochastic integral in $Z_t^x(W)$, equation (3.20) becomes:

$$\begin{aligned} \frac{d\mathbb{P}_\theta^{t,x,y}}{d\mathbb{W}^{t,x,y}} &= \frac{\mathcal{N}(y-x,t)}{\tilde{p}(t,x,y;\theta)} \\ &\times \exp\left\{ A(y;\theta) - A(x;\theta) - \frac{1}{2}\int_0^t \left[\alpha^2(W_s;\theta) + \alpha'(W_s;\theta)\right] ds \right\}. \end{aligned} \tag{3.21}$$

Finally, taking expectation of (3.21) w.r.t. the bridge measure $\mathbb{W}^{t,x,y}$ provides:

$$\begin{aligned} \tilde{p}(t,x,y;\theta) &= \mathcal{N}(y-x,t)e^{\{A(y;\theta)-A(x;\theta)\}} \\ &\times E_{\mathbb{W}^{t,x,y}}\left[\exp\left\{-\frac{1}{2}\int_0^t \left[\alpha^2(W_s;\theta)+\alpha'(W_s;\theta)\right]ds\right\}\right]. \end{aligned} \tag{3.22}$$

Formula (3.22) is crucial in many inferential approaches for discretely observed diffusions. In general, the expected value in (3.22) is intractable, but it can be estimated via numerical methods. In particular, a Monte Carlo approach would be repeatedly drawing from the measure $\mathbb{W}^{t,x,y}$ and averaging over the obtained realizations of the functionals on the r.h.s of (3.22). In practice, of course, one has to approximate the integral in (3.22) with a finite sum, i.e., to discretize the integral. This could imply bias that can be reduced by taking a discretization grid fine enough. Moreover, as we will see in the next Section, for a class of diffusions, an "unbiased " estimator of (3.22) can be obtained in the spirit of the restrospective sampling.

3.5.4 The acceptance method

In this section we describe, following Beskos, Papaspiliopoulous and Roberts (2006), how the idea of retrospective sampling can be used to obtain an unbiased estimator of the transition density $\tilde{p}(t,x,y;\theta)$ of the process X with unit volatility. Suppose that, for a given value of θ, the drift of the transformed process $\alpha(x,\theta)$ satisfies the boundness conditions required in order for exact algorithm to be applicable (see Chapter 2). In this case, it has been shown in Chapter 2, Section 2.2.6, that an exact algorithm for the simulation of conditional diffusions (diffusion bridges) can be obtained according to an appropriate acceptance-rejection scheme. The latter is based on accepting the proposed path on $[0,t]$ with probability:

$$P_a(W;\theta) = \exp\left\{-\int_0^t \Phi(W_s;\theta)ds\right\}, \tag{3.23}$$

or

$$P_a(W;\theta) = \exp\left\{-r(W;\theta)\int_0^t \Phi(W_s;\theta)ds\right\}, \tag{3.24}$$

depending on whether definition (2.13 or 2.32 Chapter 2]) is used for Φ (note that in (3.23) and (3.24), the same notation as in Capther 1 is used except for explicit dependence on parameter θ). Comparison with (3.22) shows that the target transition density can be related to the *unconditional* acceptance probability of the exact algorithm via the relation:

$$\tilde{p}(t,x,y;\theta) = \mathcal{N}(y-x,t)e^{\{A(y;\theta)-A(x;\theta)-l(\theta)t\}}a(x,y,\theta), \tag{3.25}$$

where

$$a(x,y,\theta) = E_{\mathbb{W}^{t,x,y}}\left[P_a(W;\theta)\right]. \tag{3.26}$$

Explicit computation of the expected value (3.26) is not feasible, but a Monte Carlo approximation of (3.26) can be obtained by means of the *unconditional acceptance frequency* of the proposed paths in the EA, over a large number of proposals from the measure $\mathbb{W}^{t,x,y}$. Thus, an unbiased estimator of $\tilde{p}(t,x,y;\theta)$ can be obtained by plugging-in the MC approximation of $a(x,y,\theta)$ in the expression (3.25). To be concrete, assume that, using the same notation of chapter 1, Section 2.6, Φ is defined as $\frac{1}{r(W)}\left\{\frac{\alpha^2(u;\theta)+\alpha'(u;\theta)}{2} - l(\theta)\right\}$, so that, given a proposal W (path), the conditional acceptance probability is given by (3.24). In this case, in the retrospective sampling framework, an auxiliary planar Poisson process Ψ on $[0,t]\times[0,1]$ is used with intensity $r(W)$ (see Chapter 2). Specifically, realizations from Ψ are obtained by drawing an integer $\mathcal{K}$ from a Poisson distribution with parameter $r(W)\cdot t$, then, conditional on $\mathcal{K}$, drawing $\mathcal{K}$ points $(\tau_1,\psi_1),\ldots,(\tau_{\mathcal{K}},\psi_{\mathcal{K}})$ from the product measure $\mathbb{U}(0,t)\otimes\mathbb{U}(0,1)$ on $\{[0,t]\times[0,1]\}$ where $\mathbb{U}(a,b)$ denotes the law of a uniform r.v. on $[a,b]$. We recall that the EA is based on comparing the values $(\psi_1,\ldots,\psi_{\mathcal{K}})$ with the values taken by the function Φ on the proposed skeleton S. The latter is obtained by randomly drawning values $(W_{\tau_1},\ldots,W_{\tau_{\mathcal{K}}})$ of the Brownian bridge $BB_{(x,0)\to(y,t)}$ at times $(\tau_1,\ldots,\tau_{\mathcal{K}})$. Thus, the acceptance indicator in the EA can be expressed as:

$$I(x,y,\theta,\Psi,S) = \prod_{i=1}^{\mathcal{K}} \mathbb{I}\left\{\Phi(W_{\tau_i};\theta)\leq\psi_i\right\}, \tag{3.27}$$

where $\mathbb{I}$ is the indicator function. Thus, if M proposals (skeletons) $S^j, (j=$

$\ldots, M)$ are generated according to the EA, an unbiased estimator $\hat{a}(x, y, \theta)$ of $a(x, y, \theta)$ can be obtained as:

$$\frac{\sum_{j=1}^{M} I(x, y, \theta, S^j)}{M}.$$

Correspondigly, an unbiased estimator of $\tilde{p}(t, x, y; \theta)$ is:

$$\hat{\tilde{p}}(t, x, y; \theta) = \mathcal{N}(y - x, t) e^{\{A(y;\theta) - A(x;\theta) - l(\theta)t\}} \hat{a}(x, y, \theta). \tag{3.28}$$

As remarked in Beskos, Papaspiliopoulous and Roberts (2006), the estimator (3.28) has the nice property of being almost surely bounded, so that all its moments are finite.

3.5.5 Simultaneous estimation methods

It is important noting that methodologies based on MC estimates of the transition density for *a given value* of the parameter θ provide independent estimates for different values of this parameter. This can result in inconsistent estimators of some characteristics of the likelihood function such as the maximum. In order to guarantee consistency of the maximum likelihood estimator, one should estimate the *likelihood function itself* rather than the particular values that it takes on some grid of the parameter space Θ. This problem is related to the convergence of random variables in Banach spaces. In particular, suppose that for each pair of consecutive points $(X_{t_i}, X_{t_{i+1}})$, $i = 0, \ldots, n$ (possibly obtained from the observations through the transformation (3.15)) we are able to use a Monte Carlo procedure whose output

$$\tilde{p}_i^N(\xi_1^i, \ldots, \xi_N^i, \cdot) = \frac{1}{N} \sum_{j=1}^{N} \mathcal{P}(\xi_j^i, \cdot),$$

is an average of bounded *random functions* $\theta \to \mathcal{P}(\xi_j^i, \theta)$ continuous in θ and depending on *iid* random elements $\xi_1^i, \ldots, \xi_N^i$ which in turn are independent of θ. Then, we can appeal to the strong law of large numbers in Banach spaces to establish that the random sequence $\tilde{p}_i^N(\xi_1^i, \ldots, \xi_N^i, \theta)$ converges a.s. uniformly in θ to $E[\mathcal{P}(\xi_1^i, \theta)]$ as $N \to \infty$ (here the parameter space Θ is assumed to be compact). Moreover, if for $j = 1, \ldots, N$, $\mathcal{P}(\xi_j^i, \theta)$ is an unbiased estimator of the transition density $\tilde{p}(\Delta_{t_i}, X_{t_{i-1}}, X_{t_i}; \theta)$, then

$\tilde{p}_i^N(\xi_1^i, \ldots, \xi_N^i, \theta) \to \tilde{p}(\Delta_{t_i}, X_{t_{i-1}}, X_{t_i}; \theta)$ a.s. uniformly in θ. Uniform convergence of the single terms corresponding to pairs of consecutive observations implies uniform convergence of the estimator:

$$L_n^N(\theta) = \prod_{i=1}^{n} \tilde{p}_i^N(\xi_1^i, \ldots, \xi_N^i, \theta),$$

to the true likelihood:

$$L_n(\theta) = \prod_{i=1}^{n} \tilde{p}(\Delta_{t_i}, X_{t_{i-1}}, X_{t_i}; \theta),$$

based on the points $X_{t_0}, \ldots X_{t_n}$. This in turn implies that the estimator

$$\hat{\theta}_n^N = \arg\max_{\theta \in \Theta} L_n^N(\theta)$$

converges a.s. to the maximum likelihood estimator:

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(\theta).$$

Here the crucial point is the a.s. uniform convergence of the MC approximations of the transition density. This is because uniform convergence ensures that the estimator obtained by maximizing the approximate likelihood converges a.s. to true maximum likelihood estimator. As mentioned at the beginning of this Section, to establish uniform convergence, one has to use results from the theory of random variables taking values in separable Banach spaces. In fact, under the previous hypoteses, the approximating functions $\mathcal{P}(\xi^i, \cdot)$ take values in the Banach space $(C(\Theta), \| \ \|_\infty)$ of the continuous functions on Θ with the uniform norm:

$$\|f\|_\infty = \sup_{\theta \in \Theta} f(\theta).$$

Thus, uniform convergence of the likelihood approximating functions is the same as convergence in norm in a Banach space. It is well known (D.P.Giesy, 1976) that, if a sequence of *iid* r.v.s $f_1, \ldots, f_N$ in a Banach space $\mathcal{B}$ with norm $\|\cdot\|$ satisfies $E[\|f_j\|] < \infty$ and $E[f_j] = 0$, for $j = 1, \ldots N$, then, a.s.:

$$\lim_{N \to \infty} \left\| \frac{1}{N} \sum_{j=1}^{N} f_j \right\| = 0.$$

Note that for a typical element f of the Banach space $\mathcal{B}$, $E(f)$ is defined as the only vector $\xi \in \mathcal{B}$, such that $E[T(f)] = T(\xi)$ for each linear functional T in the dual space of $\mathcal{B}$.

The desired convergence is obtained by taking

$$f_j(\cdot) = \mathcal{P}(\xi_j^i; \cdot) - \tilde{p}(\Delta_{t_i}, X_{t_{i-1}}, X_{t_i}; \cdot).$$

In the last part of this section, we will show how dependence on θ can be made explicit in the random function that approximates the transition density corresponding to values x and y at times 0 and t respectively. In the case of the acceptance estimator, in Beskos, Papaspiliopoulous and Roberts (2006) and Beskos et al. (2009) methods are illustrated for EA1 and EA2 respectively. In both cases, the random elements required by the algorithm for MC estimation have to be expressed as deterministic functions of θ and of random variables independent of the parameters. To this aim, specific properties of the various random elements are exploited. Let us first describe the approach for EA1. In this case, when $\tilde{p}(t, x, y)$ is to be estimated, the relevant random elements are the auxiliary Poisson process and the Brownian bridge between x and y. Since the functional Φ in (3.23) is bounded, for a fixed value of θ, the intensity of the Poisson process can be taken independent of the proposed path (differently from the EA2 case), i.e., $r \equiv r(\theta)$. Assume that, as θ varies in the parameter space Θ, we can find an upper bound r_M such that $r < r_M$. Then, using the thinning property of the Poisson process, we can obtain random draws from the Poisson process Ψ_M with intensity $r(\theta)$ on the rectangular area $R \doteq [0, t] \times [0, 1]$ by first drawing points on R from a Poisson process with intensity r_M, and then dropping each generated point with probability $1 - r(\theta)/r_M$. In practice, we draw a number from a Poisson r.v. $\mathcal{K}$ with parameter r_M, and generate from a $\mathcal{K}$-variate random vector $\tau \equiv (\tau_1, \ldots, \tau_{\mathcal{K}})$ whose components are uniform r.v.s on $[0, t]$. Then, we draw from $\mathcal{K}$ independent r.v.s $(\psi_1, \ldots, \psi_{\mathcal{K}}) \equiv \psi$ uniform on $[0, 1]$ and drop each point (τ_j, ψ_j) in R with probability $1 - r(\theta)/r_M$, i.e., if $u_j > r(\theta)/r_M$, where $u = \{u_j\}_{(j=1,\ldots,\mathcal{K})}$ are other independent random variables uniformly distributed on $[0, 1]$. In order to eliminate dependence on parameter θ from the random elements needed for the acceptance algorithm, it remains to express the Brownian Bridge $BB_{(x,0)\to(y,t)}$ between the θ-dependent points x and y in terms of a deterministic transformation of a standard Brownian bridge $BB_{(0,0)\to(0,t)}$. To this aim, we note that, from the standard properties of the Brownian bridge, it follows that a sample path from the bridge measure $\mathbb{W}^{t,x,y}$ can be obtained from a sample path W from the the bridge measure $\mathbb{W}^{t,0,0}$ via the transformation:

$$W_s \to W_s + (1 - s/t)x + (s/t)y, \tag{3.29}$$

It follows that the indicator variable (3.27) for the acceptance condition can be re-written as:

$$I(x, y, \theta, \Psi_M, u, S) = I(x, y, \theta, \tau, \psi, u, S) \tag{3.30}$$

$$= \prod_{j=1}^{\mathcal{K}} \mathbb{I}\left\{\mathbb{I}\left\{u_j < \frac{r(\theta)}{r_M}\right\} \Phi\left[W_{\tau_i} + (1 - \tau_j/t)x + (\tau_j/t)y; \theta\right] \leq \psi_j\right\},$$

i.e., the random elements $\{u_j\}_{(j=1,\ldots,\mathcal{K})}, \{\tau_j\}_{(j=1,\ldots,\mathcal{K})}, \{\psi_j\}_{(j=1,\ldots,\mathcal{K})}$, are all independent of θ. Taking expectation of (3.30) with respect to the distribution of the uniform random variates u, we obtain:

$$E\left[I(x, y, \theta, u, \psi, S) | \, \tau, S\right] \tag{3.31}$$

$$= \prod_{j=1}^{\mathcal{K}} \left\{\left[1 - \frac{r(\theta)}{r_M}\right] + \frac{r(\theta)}{r_M} \Phi\left[W_{\tau_i} + (1 - \tau_j/t)x + (\tau_j/t)y; \theta\right]\right\}$$

$$= \prod_{j=1}^{\mathcal{K}} \left\{1 - \frac{r(\theta)}{r_M} \Phi\left[W_{\tau_i} + (1 - \tau_j/t)x + (\tau_j/t)y; \theta\right]\right\}.$$

Thus, one can obtain an unbiased *simultaneous* estimator of (3.26), and then of the density function (3.25) by averaging over a large number of realizations from the joint probability distribution of the Poisson process on $[0, t]$ with intensity r_M and the standard Brownian bridge on $[0, t]$.
In the case of the EA2 algorithm the simultaneous acceptance method (SAM) is much more complicated. In this case the intensity of the Poisson process used to decide if a proposed path is to be accepted, depends on some characteristics of the path itself, namely the minimum m and the time t_m at which

it is attained. These last random elements have to be generated from the appropriate probability distributions which depend, in turn, on the model parameters. However, it is not difficult to show that, as in case of EA1 algorithm, dependence on parameters can be decoupled from the relevant probability distributions. For instance, a realization of the minimum of the Brownian bridge between two given points can be obtained through a deterministic function of a realization from an exponential r.v. with unit mean (see, Section 2.2.4). Analogously, random drawing the time t_m at which the minimum m is attained involves drawing from an Inverse Gaussian distribution with suitable parameters. Again, one can obtain realizations from this distribution by random drawing from a squared standard normal r.v. and using the relation (2.27) of Section 2.2.4 As a matter of fact, the simultaneous acceptance estimator based on algorithm EA2 requires drawing the random elements $\{E, Z, \Psi, N\}$, where E is exponential with unit mean, Z is a standard Gaussian variable, Ψ are Λ points from a homogeneous Poisson process on $[0, t]$ with parameter of intensity λ which, given E (and hence m), is independent on θ, and N is a collection of $3 \times \Lambda$ standard normal variables. Details are provided in Beskos et al. (2009), where the statistical properties of the SAM estimator are also illustrated.

Starting from formula (3.22), simultaneous acceptance methods can be used also when exact algorithm (and hence SAM estimator) is not feasible. In fact, all one needs is using a parameter-independent procedure to approximate the integral in (3.22) with a discrete sum and to take the MC average over the realizations of the corresponding functional. To this aim, it sufficies to obtain realizations from $\mathbb{W}^{t,x,y}$ by random drawing from the parameter-independent measure $\mathbb{W}^{t,0,0}$ and then using the transformation (3.29). More precisely, once a grid $H \equiv (0 = t_0, t_1, \ldots, t_h = t)$ has been chosen fine enough to well approximate the integral, at the jth MC iteration ($j = 1, \ldots, N$), $h-1$ values $W^0_{j1}, W^0_{j2}, \ldots, W^0_{jh}$ are drawn from the finite-dimensional distribution of the standard Brownian bridge (i.e., starting and ending in 0) corresponding to time points in H. Thus, if according to the notation of Section 3.5.2, $x = \eta(u; \theta),\ y = \eta(v; \theta)$ the MC approximation is:

$$\theta \to \frac{1}{N} \sum_{j=1}^{N} \exp\left\{ -\sum_{k=1}^{h} \frac{\alpha^2 + \alpha'}{2} \left[W^0_{jk} + \left(1 - \frac{t_k}{t}\right) x + \frac{t_k}{t} y; \theta \right] \right\}.$$

Correspondingly, we have the following approximation of the target transition density:

$$p(t,u,w;\theta) = \tilde{p}(t,x,y;\theta)|\eta'(v;\theta)| \tag{3.32}$$

$$\begin{aligned}
&\approx \mathcal{N}(y-x,t)\exp\left\{A(y;\theta)-A(x;\theta)\right\} \\
&\times \frac{1}{N}\sum_{j=1}^{N}\exp\left\{-\sum_{k=1}^{h}\frac{\alpha^2+\alpha'}{2}\left[W_{jk}^0+\left(1-\frac{t_k}{t}\right)x+\frac{t_k}{t}y;\theta\right]\right\} \\
&\times |\eta'(v;\theta)|.
\end{aligned}$$

3.6 Missing data approach

A common approach for parametric inference on discretely observed diffusion processes is based on considering the continuous paths between consecutive observations as *missing data*. In this framework, inference is often based on recovering unobserved path in suitable sense. In some contexts, the problem of re-constructing "missing data" between available observations is referred to as an *imputation* problem. Since for a discretely observed diffusion process the observed data are process values measured at a finite collection of times, re-construction of missing information involves imputation of paths between pairs of consecutive observations. This requires managing the distribution law of the process conditional on taking certain values at the extremes of a time interval (diffusion bridge). Since in most cases the latter distribution is not explicitly available, numerical techniques are generally used to draw realizations from the target diffusion bridge law. In contrast to the case of unconditional simulation, where Eulerian-Marujana schemes can be successfully applied (provided that a sufficiently fine discretization is used), simulation of diffusion bridges is a highly non trivial problem. However, as noticed in section 2.2.6, for a class of diffusions, the conditional version of the restrospective sampling algorithm can be easily used to generate paths with given values at the extremes of a time interval, from the "exact" law of the corresponding diffusion bridge.

3.6.1 Expectation Maximitazion algorithm

Among the inferential methods based on the missing data approach, an important role is played by likelihood maximization via expectation-maximization (EM) algorithm (Dempster et al., 1977). This is based on repeatedly applying, until convergence, an expectation step (E-step) and a maximization step (M-step). Loosely speaking, the E-step computes the expectation of the *complete data log-likelihood* with respect to the distribution of the missing data conditional on the observed data and the current values of the model parameters. The M-step maximizes the outcome of the E-step as function of the parameters. The algorithm produces a sequence of parameters that converge, under some regularity conditions, to the (in general local) maximum of the *observed data log-likelihood*. In the present context, the definition of complete data likelihood is a subtle point. In fact, it crucially depends on whether the diffusion coefficient contains or not unknown parameters. In the latter case, one can use the Girsanov theorem to define the complete data likelihood as the continuous time log-likelihood (see Section 3.3). Specifically, let us assume the model:

$$dY_t = b(Y_t, \theta)dt + \sigma(Y_t)dW_t, \quad Y(0) = y_0, \tag{3.33}$$

where the parameter θ in the drift is to be estimated. Without loss of generality, consider the case where the observed data are just the two consecutive observations $Y_0 = u$ and $Y_t = v$.

In order to make EM algorithm feasible, we transform the process Y in a new process X with unit volatility using transformation (3.15) of Section 3.5.2. Note, however, that in the present case, the function $\eta(\cdot)$ does not contain unknown parameters, and consequently the "transformed data-points" $x \equiv \eta(u),\ y \equiv \eta(v)$ can be considered as really observed. As illustrated in Section 3.5.2, the new process is solution of the SDE (3.16) with drift function:

$$\alpha(u, \theta) = \frac{b(\eta^{-1}(u), \theta)}{\sigma(\eta^{-1}(u))} - \frac{1}{2}\sigma'(\eta^{-1}(u)).$$

Using the Girsanov theorem as in Section 3.5.3, it is easy to show that the complete-data log-likelihood l_c, based on the path $\{X_s;\ 0 \le s \le t\}$ can be written as

$$l_c = A(y; \theta) - A(x; \theta) - \frac{1}{2}\int_0^t \left[\alpha^2(X_s; \theta) + \alpha'(X_s; \theta)\right] ds, \tag{3.34}$$

where

$$A(w;\theta) = \int_c^w \alpha(z;\theta)dz,$$

with c arbitrary numerical constant.
Given the current value of the parameter θ, the E-step consists in taking expectation of (3.34) with respect to the law of the target diffusion X conditioned on $\{X_0 = x,\ Y_0 = y\}$, i.e., with respect to the diffusion bridge measure $\mathbb{P}^{t,x,y}$. Typically the latter is analytically intractable, nevertheless in situations where exact algorithm is feasible, one can easily approximate the expectation of the integral in (3.34) through an MC average based on exact simulations from $\mathbb{P}^{t,x,y}$. Beskos, Papaspiliopoulous and Roberts (2006) suggest to draw values from a uniform random variable $\mathcal{U}$ on $[0,t]$, and for each realized value U using the conditioned version of EA to draw from the distribution of the univariate r.v. X_U. An approximation of the E-step can be obtained by averaging over independent realizations of the r.v.:

$$A(y;\theta) - A(x;\theta) - \frac{t}{2}\left[\alpha^2(X_U;\theta) + \alpha'(X_U;\theta)\right]. \tag{3.35}$$

The algorithm is an example of Monte Carlo Expectation Maximization algorithm (MCEM) inroduced by Wei and Tanner (1990). Convergence properties and practical implementation issues can be found in Chan and Ledolter (1995), Sherman et al. (1999), and Fort and Moulines (2003).

The case when the diffusion function depends on some unknown parameters, i.e., $\sigma \equiv \sigma(y,\theta)$ is more involved. In fact, in this case, considering the unobserved path between consecutive observations as missing data results in a data augmentations scheme where, according to the terminology of the incomplete data literature, the fraction of missing information is 1. This is because, due to the relation between diffusion function and quadratic variation of the process, the knowledge of the continuous path between pair of observations allows *perfect estimation* of the volatility function. In order to make the EM algorithm feasible, a re-parametrization of the process is needed. The re-parametrization consists of two steps. The first step is the same as in the case where the diffusion term is parameter-independent and is simply the definition of the process X with unit volatility. The aim of the second step is to re-define the process so that the missing data are identified with the path of a suitable diffusion bridge starting and ending in 0. The need of this second step arises from the fact that in the present case, the

transformation that defines the unit volatility process depends on the parameters, and consequently the transformed data points $x \equiv \eta(u;\theta),\ y \equiv \eta(v;\theta)$ are no longer directly observed. In detail, let X be solution of the SDE:

$$dX_s = \alpha(X_s,\theta)dt + dW_s,$$

where α is obtained from the drift in the original SDE through the formula (3.17). Let Z be the new process defined by:

$$Z_s \doteq X_s - \left(1 - \frac{s}{t}\right) x - \frac{s}{t} y, \qquad s \in [0,t]. \tag{3.36}$$

The process Z is a diffusion bridge taking values 0 at the extremes of the interval $[0,t]$. For a given value of the parameter θ, it is possible to draw paths from the law of Z by first drawing via EA from the bridge measure $\mathbb{P}^{t,x,y}$ induced by the conditioned process X, and then using (3.36). The unobserved path of Z between 0 and t can be considered as "missing data" in a data augmentation scheme where the observed data are: $\{Y_0 = u, Y_t = v\}$. In this framework, the complete data likelihood is:

$$\begin{aligned} l_c \;=\; & \log|\eta'(v;\theta)| + \log[\mathcal{N}(y-x)] + A(y;\theta) - A(x;\theta) \\ & - \int_0^t \frac{1}{2}\left\{\alpha^2(g_\theta(Z_s);\theta) + \alpha'(g_\theta(Z_s);\theta)\right\} ds, \end{aligned} \tag{3.37}$$

where $g_\theta(\cdot)$ is the inverse transformation of (3.36), namely:

$$g_\theta(Z_s) \doteq Z_s + \left(1 - \frac{s}{t}\right) x + \frac{s}{t} y, \qquad s \in [0,t].$$

As in case of parameter independent volatility, expectation of (3.37), given $\{Y_0 = u, Y_t = v\}$ and the current value of θ, can be approximated via MC average based on random draws from the Z law. Thus, using the same approach as in the previous case, the E-step can be approximated by averaging independent realizations of the r.v.:

$$\log|\eta'(v;\theta)| + \log[\mathcal{N}(y-x)] + A(y;\theta) - A(x;\theta)$$

$$-\frac{t}{2}\left\{\alpha^2(g_\theta(Z_U);\theta) + \alpha'(g_\theta(Z_U);\theta)\right\} ds$$

where U is a uniform r.v. on $[0, t]$.

3.6.2 Bayesian Inference

As in the case of classical analysis, also Bayesian inference on discretely observed diffusion processes is problematic for diffusion whose transition density is unavailable. Occasionally, when the time distances between consecutive observations is sufficiently small, one can base the analysis on a suitable approximation of the continuous time likelihood (Polson and Roberts, 1994). However, in cases where the diffusion process is not observed at "high frequency", discretization of the integral in the continuous time likelihood (see Section 3.3) can result in strong bias. In such cases, a better approach can rely on a data augmentation scheme (Tanner and Wong, 1987). The latter is a Gibbs sampling scheme based on alternately imputing missing data (data augmentation) conditionally on the current values of the parameters, and drawing parameters from the posterior distribution defined in terms of the likelihood based on complete data (i.e., including currently imputed data). This scheme produces a Markov chain which, under suitable conditions, has the joint posterior distribution of missing data and parameters as its invariant distribution. In this context, as illustrated in the previous Subsection, the missing data are the unobserved path between observations. Thus, due to the Markov property, imputation of missing data reduces to drawing independently from the appropriate diffusion bridge law. In general, the performance of the data augmentation algorithm, in terms of convergence properties, crucially depends on the fraction of missing information available in the observed data, with respect to the information contained in the complete data. This relation is strictly related to the one between missing information and speed of convergence of the EM algorithm in the approach based on likelihood maximization. The problem of data augmentation for diffusion processes is that the fraction of missing information associated with parameter in the diffusion function is 1. This phenomenon directly depends on the fact that, due to the relation (3.3) between quadratic variation and diffusion function in the Ito's processes (see Section 3.2), the knowledge of the continuous path on a (although finite) time interval allows "perfect estimate" of the diffusion term. An implication is that, when Markov Chain Monte Carlo (MCMC) schemes are used in the context of Bayesian inference, the better one tries to approximate the continuous unobserved path (i.e., the higher is the amount of augmented information), the worse the algorithm becomes (Roberts and Stramer, 2001). To better understand this problem, consider the case where

the diffusion coefficient is just a multiplicative parameter, i.e., $\sigma(\cdot, \theta) \equiv \sigma$. Then, in the limit case corresponding to imputation of a continuous path between any pair of consecutive observations, the posterior distribution of σ^2 given the "completed information" is just a mass point on the current value of σ^2, so that drawing from this distribution just confirms this value. In other words, the chain is reducible. In order to overcome this difficulty, Roberts and Stramer (2001) use essentially the same strategy illustrated in Subsection 3.6.1 consisting of a transformation of the diffusion into a new process with unit volatility and a re-parametrization. The first step allows to remove dependence on unknown parameters from the diffusion coefficient, so that the continuous likelihood for the new unit volatility process can be expressed via the Girsanov formula using a dominating measure independent of the parameter (namely, the Wiener measure). As in the case of EM algorithm, the aim of the re-parametrization (see formula (3.36) in Subsection 3.6.1), is to define a data augmentation scheme where the missing data are the continuous path of a diffusion bridge starting and terminating at 0. In Roberts and Stramer (2001) a MCMC strategy is presented where, given a prior distribution for the parameters, a two-step Metropolis-Hastings (MH) algorithm is used which (after a suitable burn-in period) alternates drawing from the posterior parameter distribution given the current diffusion trajectory, and drawing from the law of the diffusion bridge resulting from the re-parametrization. In this second step, a Brownian bridge, or a suitable conditioned Ornstein-Uhlenbeck process can serve as diffusion bridge to be used as proposal in the MH algorithm.

In situations where EA is feasible for generating trajectories from the law of the (suitably transformed) target diffusion bridge, the second MH step is unnecessary. Rather, at each iteration of the MCMC algorithm, it is sufficient to use, as augmented information, the skeletons of the diffusion bridge produced by the (conditioned) EA. In fact, in Beskos, Papaspiliopoulous and Roberts (2006), a hierarchical MCMC scheme is presented where, given the current elements obtained as output of the EA (i.e., the random elements of the auxiliary Poisson process used to decide if the proposed path is to be accepted and the values of the process at the corresponding random times, see Chapter 2, Section 2.2), and given the observed values of the process, the parameters are independent of the entire unobserved path. Thus, this approach is not directly based on augmentation of paths via imputation of a large number of process values, but only requires simulating the process at a few (random) times, and drawing the parameter from their posterior distribution conditioned on the realized skeletons. As a consequence, the state space of the MCMC algorithm has a dimension much smaller than with traditional data augmenttion scheme, resulting in a higher computational efficiency.

Chapter 4

Simulation and inference for the stochastic Bass model

4.1 Simulation of trajectories from the SBM process

We have seen that some characteristics of the SBM, such as ($a.s.$) regularity on $(0, K)$, ergodicity, or stochastic stability of the trivial solution $Y_t \equiv 0$, can be determined without explicitly knowing the solution of the SDE (1.5). However, since the transition density of the process is not explicitly known, analytic expressions of many characteristics of interest are not available. For instance, in marketing applications, the interest is often focused on expected number of adopters of some new products or services at a specific time, or simmetrically, on the expected time for number of adopters to become larger than some critical threshold (hitting time). Since analytic computation of these quantities requires use of conditional density, numerical approximations are necessary. The most natural way to obtain approximations for functionals (e.g., moments) of the finite dimensional distributions of the SBM process is to use Monte Carlo averages based on simulations of sample paths. As illustrated in Section 2.1, a possible simple method to obtain realizations from the solution of the SDE (1.14) on $[0, T]$ is based on the Euler-Maruyama scheme. However, beside the computational cost deriving from partitioning the interval $[0, T]$ in a high number of sub-intervals and using a Gaussian approximation on each of them, another problem arises when simulating SBM trajectories via Eulerian methodology. In fact, Euler approximation does not guarantee regularity of the solution in $[0, K]$. This can negatively affect analyses focusing on hitting time of some regions inside the interval $[0, K]$. Thus, in order to garantee regularity, one has to use some regularization

scheme like the *Balanced Implicit Method (BIM)* (Shurz, 2007), capable of adjusting the basic Eulerian procedure so that the regularity properties of the exact solution are preserved. In the particular case of no self innovation ($a = 0$), we will show that *exact* solutions of equation (1.14) can be simulated at a finite but arbitrary collection of times, by means of exact simulation based on the *retrospective sampling* technique described in Chapter 2.

4.1.1 Exact Simulation for SBM without self-innovation

Let $Y_t,\ t \in [0, T]$ be a strong solution of the SDE:

$$dY_t = \left[\frac{b}{K} Y_t(1 - Y_t) - \mu Y_t\right] dt + \frac{\sigma}{K} Y_t(1 - Y_t) dW_t \tag{4.1}$$

with initial condition $Y(0) = y_0 \in (0, K)$. This model will be referred to as SBM1. We want to obtain a realization of the n-dimensional random vector $(Y_{t_1}, \ldots, Y_{t_n})$, where $0 \leq t_1 < \ldots < t_n \leq T$ are arbitrary times. If the EA is applied on the entire interval $[0, T]$, the output of the algorithm is a set of instances of the process at a finite collection of random times $\tau_1, \ldots, \tau_{\mathcal{K}}$ ($\mathcal{K}$ random), plus the time T (right extreme of the interval). In this case, exploiting the Markov property, the values $Y_{t_i},\ (i = 1, \ldots, n)$ at times of interest can be obtained through the dynamic of a suitable bridge process. However, if the interval $[0, T]$ is not small enough to have high acceptance probability in the EA, it can be computationally convenient to partition the interval in small sub-intervals and apply the EA on each sub-interval (see Section 2.2.5). In the remainder of this Section, we will assume that the EA is directly applied on the interval $[0, T]$. Thus, according to the methodology described in Beskos, Papaspiliopoulos and Roberts (2006), first the value of the (suitably transformed) process at time T is generated, and then, a suitable diffusion bridge between 0 and the realized value of transformed process at time T is used to draw proposals in the acceptance-rejection scheme of the EA.

In order to apply the EA, we first transform the process Y into a new process with unit volatility. To this aim, according to the transformation illustrated in Section (1.2.3), we define the process

$$X_t \doteq \eta(Y_t), \tag{4.2}$$

where

$$\eta(y) \doteq -\frac{K}{\sigma}\int_{K/2}^{y}\frac{du}{u(K-u)} = -\frac{1}{\sigma}\log\frac{y}{K-y}. \tag{4.3}$$

Since, as proved in Section (1.1), Y is regular on $[0, K]$, the process X is well defined and is valued on $(-\infty, +\infty)$. It can be represented as solution of a SDE of the form:

$$dX_t = \alpha(X_t)dt + dW_t, \qquad X(0) = x_0 = \eta(y_0), \tag{4.4}$$

with

$$\begin{aligned}\alpha(x) &= D + \frac{\mu}{\sigma}e^{-\sigma x} - \sigma\frac{e^{-\sigma x}}{1+e^{-\sigma x}},\\ D &\doteq -\frac{b}{\sigma} + \frac{\mu}{\sigma} + \frac{\sigma}{2}.\end{aligned}$$

In the following, we describe exact simulation for the transformed process X. Once a sample path from the law of X is available, a sample path from the law of the original process Y can be obtained by inverting transformation (4.3), namely:

$$Y_t = \eta^{-1}(X_t) = K\frac{e^{-\sigma X_t}}{1+e^{-\sigma X_t}}$$

Application of the methodology described in Section 2.2 requires using the quantities:

$$\xi(x) \doteq \frac{\alpha^2(x) + \alpha'(x)}{2},$$

and

$$A(x) = \int_c^x \alpha(u)du, \tag{4.5}$$

where c is an arbitrary constant.

Straightforward calculation provides:

$$\begin{aligned} \xi(x) &= \frac{1}{2}\left[D + \frac{\mu}{\sigma}e^{-\sigma x} - \sigma\frac{e^{-\sigma x}}{1+e^{-\sigma x}}\right]^2 \\ &- \frac{1}{2}\left[\mu e^{-\sigma x} - \sigma^2\frac{e^{-\sigma x}}{(1+e^{-\sigma x})^2}\right], \end{aligned} \tag{4.6}$$

and

$$A(x) = Dx - \frac{\mu}{\sigma^2}e^{-\sigma x} + \ln\left(1+e^{-\sigma x}\right), \tag{4.7}$$

with a suitable choice of the constant c in (4.5).

It is easy to verify that the function ξ is bounded below. Thus, it is possible to find a lower bound l for ξ such that the function $\Phi = \xi - l$ to be used for exact simulation is non-negative. Since $\xi(x)$ depends on x only through the function $u(x) \doteq \exp(-\sigma x)$, it is sufficient to find a lower bound for the function

$$\begin{aligned} g(u) &\doteq \frac{1}{2}\left[D + \frac{\mu}{\sigma}u - \sigma\frac{u}{1+u}\right]^2 - \frac{1}{2}\left[\mu u - \sigma^2\frac{u}{(1+u)^2}\right]; \\ & u > 0. \end{aligned} \tag{4.8}$$

In order to find a lower bound, consider the following inequalities:

$$\begin{aligned} 2g(u) &\geq D^2 + \left(\frac{\mu}{\sigma}u\right)^2 + \frac{2D\mu}{\sigma}u - 2D\sigma\frac{u}{1+u} - 2\mu\frac{u^2}{1+u} - \mu u \\ &\geq D^2 + \left(\frac{\mu}{\sigma}u\right)^2 + \frac{2D\mu}{\sigma}u - 2\sigma\left(\frac{\mu}{\sigma} - \frac{b}{\sigma} + \frac{\sigma}{2}\right)\frac{u}{1+u} - 3\mu u \\ &\geq \left(\frac{\mu}{\sigma}\right)^2 u^2 + \left(\frac{2D\mu}{\sigma} - 3\mu\right)u + (D^2 - 2\mu - \sigma^2). \end{aligned}$$

For $u \geq 0$, the expression in the last row of the previous chain of inequalities attains its minimum at point $u_m = \max(0, u^*)$, where:

$$u^* = \frac{\sigma}{2\mu}(3\sigma - 2D) = \frac{\sigma^2 + b - \mu}{\mu}.$$

It follows that, for $u > 0$:

$$\begin{aligned} g(u) &> \frac{1}{2}\left\{\left(\frac{\mu}{\sigma}\right)^2 u_m^2 + \left(\frac{2D\mu}{\sigma} - 3\mu\right) u_m + (D^2 - 2\mu - \sigma^2)\right\} \\ &\geq \frac{1}{8}\left(4\mu - 12b - 7\sigma^2\right) \end{aligned}$$

Thus, if we put $l \doteq \frac{1}{8}\left(4\mu - 12b - 7\sigma^2\right)$, we can define the potential:

$$\Phi(x) \doteq \frac{1}{2}\left[D + \frac{\mu}{\sigma}e^{-\sigma x} - \sigma\frac{e^{-\sigma x}}{1 + e^{-\sigma x}}\right]^2 - \frac{1}{2}\left[\mu e^{-\sigma x} - \sigma^2\frac{e^{-\sigma x}}{(1 + e^{-\sigma x})^2}\right] - l,$$

Unfortunatly, the potential $\Phi(x)$ is not bounded above, hence the EA cannot be applied in its simplest version (EA1), because a region containing with certainty the graph of the function $t \to \Phi(W_t)$ does not exist, if the process W_t is allowed to vary in $(-\infty, +\infty)$. However, it is easy to check that:

$$\lim_{x \to +\infty} \sup \Phi(x) < \infty,$$

so that we can use the exact simulation in the version EA2 of Beskos, Papaspilioupolos and Roberts (2006). We refer to Section 2.2.4 for the general illustration of EA2. Here we decribe its implementation for the special case of the SBM1. Let us assume that we want to draw a sample path in $[0, T]$ from the measure $\mathbb{P}^{x_0}$ associated with the diffusion defined via the SDE (4.4). The first step is to draw the value x_T corresponding to the right extreme of the interval from the density:

$$h(x) = N \exp\left\{A(x) - (x - x_0)^2/2T\right\}, \quad x \in \mathbb{R}, \tag{4.9}$$

where the function $A(\cdot)$ has been defined in (4.7), and N is the normalization constant. Note that the function (4.9) is integrable, so that the constant N is well defined. In order to draw from the density (4.9), different algorithms can be used. In the numerical applications, we have used the adaptive rejection sampling (ars) algorithm implemented in the R-package *ars*. Recall that, in order to use the explicit expression of the Radon-Nykodin derivative of the target measure $\mathbb{P}^{x_0}$ with respect to the measure $\mathbb{Z}_T^{x_0}$ corresponding to the biased Brownian motion (formula (2.14)), we follow an A/R strategy based on acceptance or rejection of paths drawn from the law of the Brownian Bridge $BB_{(x_0,0)\to(x_T,T)}$. According to the general EA2 methodology illustrated in Section 2.2.4, we first have to generate from the appropriate distributions the minimum m of the process $BB_{(x_0,0)\to(x_T,T)}$ and the time θ at which it is attained (see Section 2.2.4 for the appropriate distributions and the methods to draw from them), then to find an upper bound M for the function $\Phi(\cdot)$ on $[m, +\infty)$. This will be also an upper bound for the function $t \to \Phi(W_t)$, $t \in [0, T]$. Thus, we have to find a number B such that:

$$\sup\{\Phi(x); x \geq m\} < B.$$

Remembering the definition (4.8) and using the chain of inequalities:

$$\begin{aligned}
2g(u) = \quad & D^2 + \left(\frac{\mu}{\sigma}\right)^2 u^2 + \sigma^2 \left(\frac{u}{1+u}\right)^2 + \frac{2D\mu}{\sigma} u \\
- \quad & 2D\sigma \frac{u}{1+u} - 2\mu \frac{u^2}{1+u} - \mu u + \sigma^2 \frac{u}{(1+u)^2} \\
\leq \quad & D^2 + \left(\frac{\mu u}{\sigma}\right)^2 + \sigma^2 + \frac{2D\mu u}{\sigma} + 2b \frac{u}{1+u} \\
- \quad & 2\mu \frac{u^2 - 1}{1+u} - \mu u + \sigma^2 \\
\leq \quad & \left(\frac{\mu}{\sigma}\right)^2 u^2 + \left(\frac{2D\mu}{\sigma} - 3\mu\right) u + D^2 + 2\sigma^2 + 2(b + \mu),
\end{aligned}$$

we find:

$$\Phi(x) \;\leq\; 0.5\left[\left(\frac{\mu}{\sigma}\right)^2 e^{-2\sigma x} + \left(\frac{2D\mu}{\sigma} - 3\mu\right) e^{-2\sigma x} + H\right] - l,$$

where

$$H \;\doteq\; D^2 + 2\sigma^2 + 2(b+\mu).$$

Finally, since the r.h.s. of the previous inequality is quadratic in $\exp(-\sigma x)$ and $0 < \exp(-\sigma x) \leq \exp(-\sigma m)$, for $m \leq x < +\infty$, it reachs it maximum at $y = m$ or at $x = +\infty$, thus we can define:

$$M \doteq \max\left\{0.5\left[\left(\frac{\mu}{\sigma}\right)^2 e^{-2\sigma m} + \left(\frac{2D\mu}{\sigma} - 3\mu\right) e^{-2\sigma m} + H\right] - l, \frac{H}{2} - l\right\}. \tag{4.10}$$

Note that M depends on the random element m, i.e., the minimum of the Brownian bridge used as a proposal in the EA. Thus, differently from the EA1 case, the rectangular area where the auxiliary Poisson process is simulated changes at each run of the EA (see Section 2.2.4).

In Figures 4.1 and 4.2 trajectories of SBM1 process simulated via exact simulation are shown. In each box, the green curve corresponds to the corresponding deterministic trajectory, i.e., the solution of the ODE obtained from the corresponding SDE by removing the diffusion term. The blue horizontal line corresponds to the potential number of adopters ($K = 1000$ in all examples). In both cases shown in Figure 4.1, $b - \mu - \sigma^2/2 > 0$, thus a stationary distribution exists. Moreover its first moment (red dashed line) is close to the deterministic equilibrium point (green dashed line). Note that in the case of the path represented on the top, the stochastic trajectory is not close to the deterministic one, although the volatility parameter is quite low ($\sigma = 0.04$). The qualitative behaviour of the sample paths simulated in Figure 4.2 is quite different. As in the previous cases, the parameters reported in the box on the top correspond to an ergodic process, i.e. one having an equilibrium distribution. However, the expected value of the r.v. with density corresponding to the stationary distribution is quite distant from the deterministic limit. Moreover, it can be shown that, for the given values of the parameters, the stationary distribution is bimodal. Finally, the path depicted in the box on the bottom of Figure 4.2 represents a situation where a stationary distribution does not exist. In fact, in this case $b - \mu - \sigma^2/2 = 0$. Thus, as proven

in section 1.2.3, the solution $Y \equiv 0$ is stochastically (asymptotically) stable, i.e., the trajectories tend to be trapped in a small neighborhood of 0.

From a computational point of view, it results that the algorithm becomes lower as the volatility parameter σ tends to zero (deterministic limit). In fact, for a fixed time interval $[0, T]$, as σ becomes smaller, an increasing number of proposals from the Brownian bridge measure are needed in order for a trajectory to be finally accepted. This is related to the fact that the upper bound (4.10) for the functional Φ defined in terms of the transformed process X, becomes larger as σ decreases implying a smaller acceptance probability (see section 2.2.5). However, as noticed in Chapter 2, the problem can be alleviated by breaking the time interval into smaller sub-intervals and applying the EA on each sub-interval. This allows one to reduce the number of points of the Poisson process used for the A/R scheme at each EA application, thus increasing the acceptance probability. In practical implementation of EA for Bass model, a useful strategy is breaking the simulation interval into shorter intervals whose length is proportional to σ^{-2}.

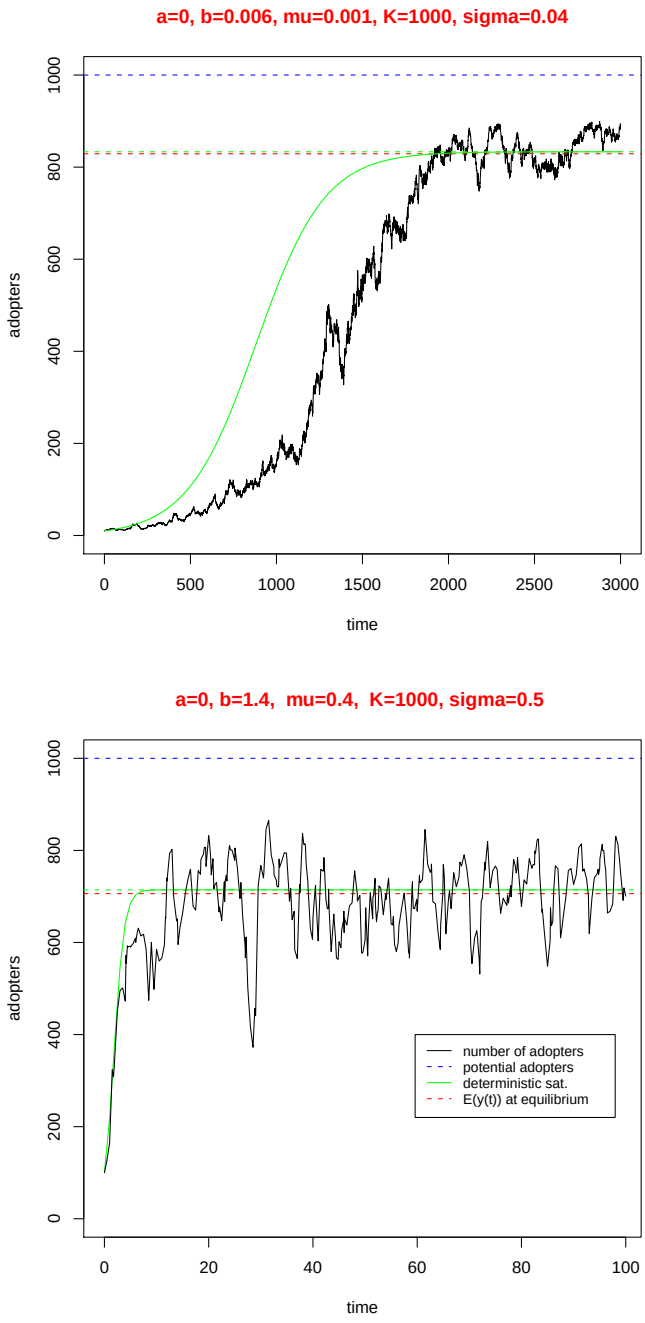

Figure 4.1: Sample paths simulated through EA2. Parameters correspond to unimodal stationary distribution.

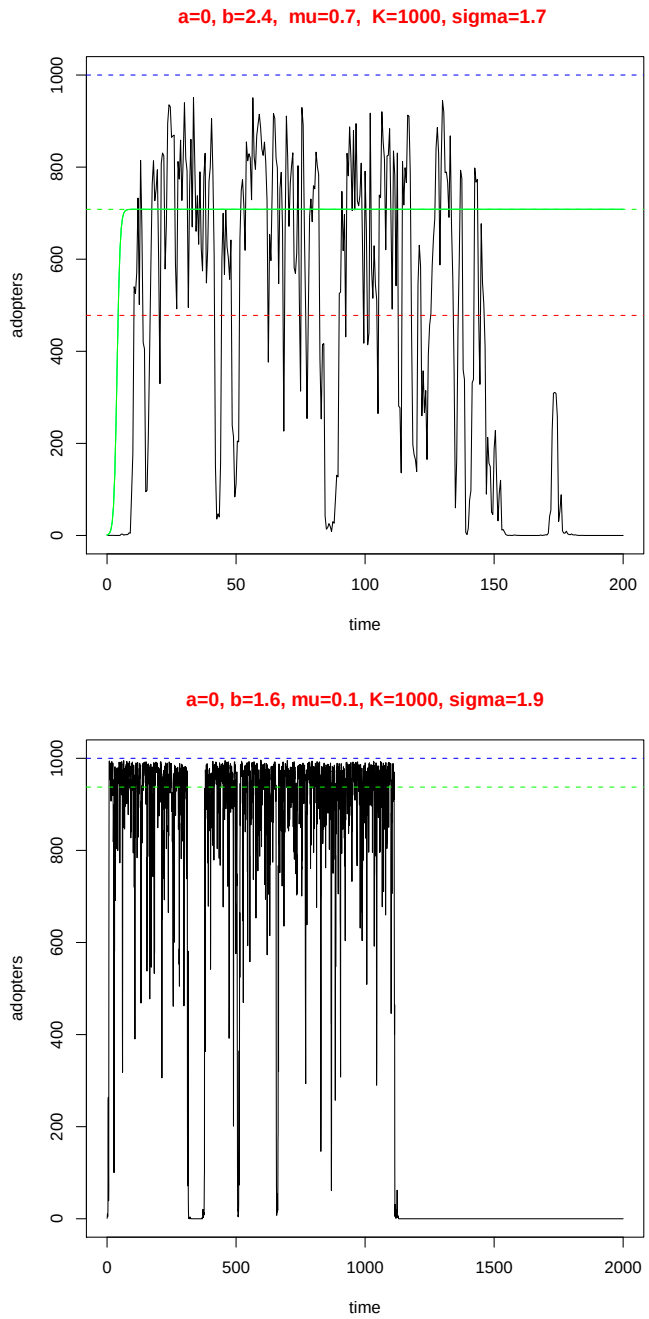

Figure 4.2: Exact simulation of SBM1 paths. Top: stationary distribution is bimodal. Down: stationary distribution does not exist.

4.2 Parameter estimation for SBM1

In this Section we deal with the problem of estimating the parameters of the Bass model without self-innovation. We recall that the parameter vector θ is composed of the following parameters: K (number of potential adopters), b (imitation coefficient), μ (disadoption rate), and σ (volatility). We will focus on the approach based on Radon Nikodyn derivative of the target diffusion measure w.r.t. the Wiener measure. In order to evaluate the methodology, a Monte Carlo study is described, where the classical approach based on the Gaussian approximation (see next subsection) of the transition density is also considered for comparison.

4.2.1 Estimation based on Gaussian approximation

In this Section we describe an estimation method of the Bass model parameters based on the Gaussian approximation of the transition density. For a general description of this approach and its relationship with the continuous likelihood discretization approach, see Section 3.3 (in particular equations (3.7)-(2.2)). Let $Y_0, \ldots, Y_n$ represent the values of the process Y at times $t_0 = 0, \ldots, t_n, (\Delta t_i = t_{i+1} - t_i)$, and approximate the dynamics of the process through the Eulerian scheme:

$$\Delta Y_i \doteq Y_{i+1} - Y_i = B(Y_i; \theta)\Delta_i + D(Y_i; \theta)\sqrt{\Delta t_i} Z_i, \tag{4.11}$$

where:

$$\begin{aligned} B(y, \theta) &= by(1-y) - \mu y, \\ D(y, \theta) &= \sigma y(1-y), \end{aligned}$$

and Z_i are standard independent Gaussian random variables ($i = 1, \ldots, n$).

Then, the corresponding approximate log-likelihood function is (up to additive constants independent of the model parameters):

$$
\begin{aligned}
L_n^G(\theta) &= -\frac{1}{2}\sum_{i=1}^{n}\log(D^2(Y_i;\theta)\Delta t_i) - \sum_{i=1}^{n}\left\{\frac{(\Delta Y_i - B(Y_i;\theta)\Delta t_i)^2}{2D^2(Y_i;\theta)\Delta t_i}\right\} \\
&= -\frac{1}{2}\sum_{i=1}^{n}(\xi_i - b\eta_i + \mu\alpha_i)^2 - \frac{n}{2}\log\sigma^2 + const \qquad (4.12)
\end{aligned}
$$

where:

$$
\xi_i = \frac{\Delta Y_i}{Y_i(1-Y_i)\sqrt{\Delta t_i}}; \quad \eta_i = \sqrt{\Delta t_i}; \quad \alpha_i = \frac{\sqrt{\Delta t_i}}{1-Y_i}.
$$

Maximizing (4.12) with respect to the parameters provides the Gaussian estimates:

$$
\hat{\mu}_g = \left[-\frac{\langle\alpha,\xi\rangle}{\langle\alpha,\alpha\rangle} + \frac{\langle\eta,\xi\rangle\langle\alpha,\eta\rangle}{\langle\alpha,\alpha\rangle\langle\eta,\eta\rangle}\right]\left[1 - \frac{\langle\alpha,\eta\rangle^2}{\langle\alpha,\alpha\rangle\langle\eta,\eta\rangle}\right]^{-1} \qquad (4.13)
$$

$$
\hat{b}_g = \left[-\frac{\langle\eta,\xi\rangle}{\langle\eta,\eta\rangle} - \frac{\langle\alpha,\xi\rangle\langle\alpha,\eta\rangle}{\langle\alpha,\alpha\rangle\langle\eta,\eta\rangle}\right]\left[1 + \frac{\langle\alpha,\eta\rangle^2}{\langle\alpha,\alpha\rangle\langle\eta,\eta\rangle}\right]^{-1} \qquad (4.14)
$$

$$
\hat{\sigma}_g = \frac{\sum_{i=1}^{n}\left(\xi_i - \hat{b}\eta_i + \hat{\mu}\alpha_i\right)^2}{n} \qquad (4.15)
$$

where we have used the notation $\langle\mathbf{u},\mathbf{v}\rangle \doteq \sum_{i=1}^{n} u_i v_i$. Note that σ could also be estimated by exploiting the relation between diffusion coefficient and quadratic variation (see the end of Chapter 3.2). In this case, if the process is observed at sufficiently high frequency, we could use the approximation:

$$
\sigma^2 \sum_{i=1}^{n} Y_i^2(1-Y_i)^2\Delta t_i \approx \sum_{i=1}^{n}\Delta Y_i^2,
$$

and derive the estimate:

$$
\hat{\hat{\sigma}} = \frac{\sum_{i=1}^{n}\Delta Y_i^2}{\sum_{i=1}^{n} Y_i^2(1-Y_i)^2\Delta t_i}. \qquad (4.16)
$$

4.2.2 MC likelihood estimation

In this Section we apply the general estimation method based on MC aproximation of the transition density described in Section 3.5.3, to the particular case of the Bass model without self-innovation. We recall that the inference on process parameters is essentially based on formula (3.22), relating the transition density to the expectation of a Brownian functional (more precisely a functional of a Brownian bridge). Actually, formula (3.22) is not applied to the original process Y, but rather to the process X obtained through the transformation (4.3). As described in Section 3.5.3, the strategy is to estimate, via Monte Carlo simulation, the transition density of the process for each pair of values observed at consecutive times. This allows one to estimate the discrete (log)likelihood function (3.9), and to maximize it with respect to the model parameters. Since the bridge simulation approach provides an estimate of the transition density for the tranformed process X, a corresponding estimate of the transition density for the original process Y must be obtained using equation (3.18). Thus, if $x = \eta(u;\theta),\ y = \eta(v;\theta)$, with

$$\eta(w;\theta) \doteq -\frac{1}{\sigma}\log\frac{w}{K-w}, \tag{4.17}$$

and $p, \tilde{p}$ are the transition functions of the processes Y and X, respectively, we have:

$$p(t,u,v;\theta) = \tilde{p}(t,x,y;\theta)|\eta'(v;\theta)| = \tilde{p}(t,x,y;\theta)\frac{K}{\sigma(K-v)}. \tag{4.18}$$

Correspondigly, the likelihood function based on a set of observations $u_0, u_1, \ldots, u_n$ at times $t_0 = 0, t_1, \ldots, t_n$ is:

$$L_n = \prod_{i=0}^{n-1} p(\Delta_i, u_i, u_{i+1};\theta) = \prod_{i=0}^{n-1} \tilde{p}(\Delta_i, x_i, x_{i+1};\theta)\frac{K}{\sigma(K-\eta^{-1}(x_{i+1}))}, \tag{4.19}$$

where $x_i = \eta(u_i,\theta)$ and $\Delta_i = t_{i+1} - t_i$.
Thus, once estimates of $\tilde{p}(\Delta_i, x_i, x_{i+1};\theta) = \tilde{p}(\Delta_i, \eta(u_i;\theta), \eta(u_{i+1};\theta);\theta)$ have been obtained for $i = 0,\ldots,n$, an estimate of L_n can be easily derived via formula (4.19). An approximate maximum likelihood (*aml*) estimate

of the model parameters θ will be obtained by maximizing the estimate of L_n. Following the approach of Section 3.5.5, the functions $\tilde{p}(\Delta_i, x_i, x_{i+1}; \theta)$, (shortly $\tilde{p}_i(\theta)$) will be estimated independently, for $i = 0, \ldots, n$. We recall that, in order the resulting approximate maximum likelihood estimator to be consistent, the transition density function has to be estimated *simultaneously* with respect to $\theta \in \Theta$. In other words, we have to estimate a *function* on the parameter space Θ, rather than to estimate independently the values it takes on a grid of points in Θ. According to the methodology illustrated in Section 3.5.5, an estimate $\tilde{p}_i^N(\theta)$ of formula (3.22) expressing the transition density $\tilde{p}_i(\theta)$ in terms of expectation of a Brownian functional, can be obtained by approximating the expectation with a MC average over N realizations of a standard Brownian bridge. A further approximation is the discretization of the integral in (3.22) through a (fine enough) partition of the time interval $[t_i, t_{i+1}]$. In practice, simultaneous MC estimation of $\tilde{p}_i(\theta)$ consists of the following steps:

MC approximation of the transition density

1. partition the interval $[0, \Delta_i]$ in m_i sub-intervals $[s_k, s_{k+1}],\ k = 0, \ldots, m_i$; $s_0 = 0, s_{m_i} = \Delta_i$.

2. draw N vectors from the m_i-dimensional distribution of a standard Brownian bridge from t_i to t_{i+1} corresponding to the times $s_k,\ k = 1, \ldots, m_i - 1$. Let W_{jk}^0 be the $N \times m_i$ matrix whose entry at row j and column k is the kth component of the jth vector.

3. put:

$$\begin{aligned} & \tilde{p}_i^N(\theta) \\ = \ & \mathcal{N}(x_{i+1} - x_i, \Delta_i) \exp\left\{A(x_{i+1}; \theta) - A(x_i; \theta)\right\} \\ \times \ & \frac{1}{N} \sum_{j=1}^{N} \exp\left\{ - \sum_{k=1}^{m_i} \xi\left(W_{jk}^0 + \left(1 - \frac{s_k}{\Delta_i}\right) x_i + \frac{s_k}{\Delta_i} x_{i+1}; \theta \right) \right\}, \end{aligned}$$

where the functions $A(\cdot)$ and $\xi(\cdot)$ have been defined in formulas (4.6) and (4.7) respectively.

Based on the n independent estimates $\tilde{p}_i^N(\theta)$ (one for each pair (u_i, u_{i+1}) of consecutive observations), a MC estimate of the likelihood function is:

$$L_n^N = \prod_{i=1}^{n} \tilde{p}_i^N(\theta) \frac{K}{\sigma(K - u_{i+1})}. \tag{4.20}$$

We highlight that the random elements necessary to obtain the matrix W^0 in p^N (i.e., standard Gaussian variables) do not explicitly depend on the parameters to be estimated. As a result, the MC approximation of the likelihood function (4.20), converges a.s. to the true likelihood function in the $\|\cdot\|_\infty$ norm. For illustration, an example of parameter estimation for SBM1 using MC approximation of the likelihood function is reported. The data are generated through the EA2 algorithm using the parameters $K = 1000, b = 0.3, \mu = 0.01, \sigma = 0.3$ (see Figure 4.3). The sample size is $n = 1000$ with observations equally spaced ($\Delta = t_{i+1} - t_i = 1;\ i = 1, \ldots, 1000$).

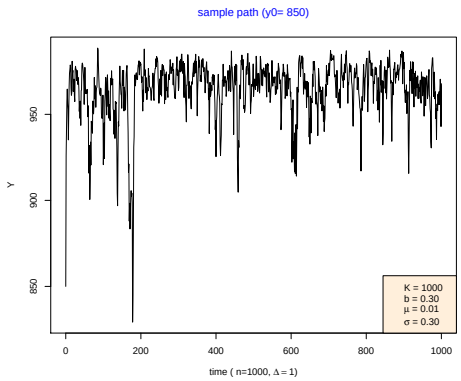

Figure 4.3: Sample path from SB1 model.

In Figure 4.4, estimated likelihood profiles are shown. The curves are obtained by computing the approximate likelihood based on $N = 100$ MC iterations. At each iteration the integral appearing in the likelihood expression has been discretized using a grid step $h = 0.1$. For each parameter, the other three are set to the corresponding true values. From the figure it results that the true values are close to the ones that maximize likelihood profile curves.

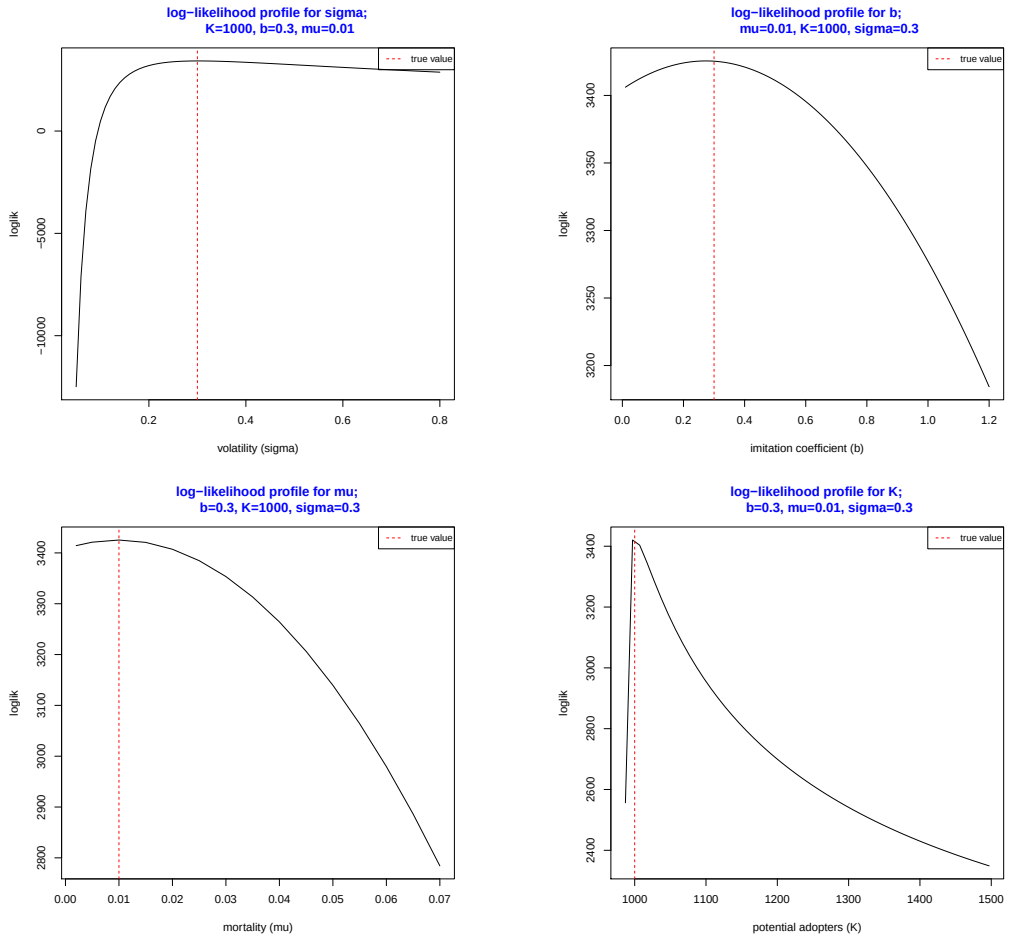

Figure 4.4: Likelihood profiles for Bass model without self-innovation.

4.2.3 Numerical evaluation

In this section, the estimation method based on Monte Carlo approximation of the transition density is evaluated through a simulation study. Preliminary estimates of the parameters are performed on a data-set containing daily number of visits to a certain web site. These "realistic" values ($\mu = 0.03, \sigma = 0.14, b = 0.32$) are used to draw from the corresponding stochastic Bass model in numerical experiments. In the experiments, we assume that K is known. This turns to be equivalent to considering the Bass model for time dynamics of proportion instead of absolute number of adopters (see equation (1.15) in Section 1.2.3). The parameter estimates are compared to the estimates obtained through the Gaussian approximation of

the transition density (see Section 4.2.1). These estimates are also used in all experiments, as starting points for the optimization routine.

The Montecarlo study is based on 1000 replicates of the following steps:

1. use EA2 algorithm to draw from the law of the n-dimensional random vector $Y_0, \ldots, Y_n$, corresponding to the values of the SBM1 process with parameter $b = 0.32$, $\mu = 0.03$, $\sigma = 0.14$, $K = 1$ at times $t_0, \ldots, t_n$, with $\Delta = t_{i+1} - t_i$, $(i = 0, \ldots, n)$ fixed. Let $u_0, \ldots, u_n$ be the corresponding realized values

2. estimate b, μ and σ via the Gaussian approximations (4.13). Let $\hat{\theta}_g = (\hat{b}_g, \hat{\mu}_g, \hat{\sigma}_g)$ be the resulting estimates.

3. apply the transformation (4.17) to the points $u_0, \ldots, u_n$. Let $x_0, \ldots, x_n$ be the corresponding transformed points

4. for each pair of consecutive points (x_i, x_{i+1}), fix a discretization step h and an integer N and apply steps (1)-(3) of the algorithm described in Section 4.2.2, with $m_i = m = int(\Delta/h) + 1$, where $int(x)$ is the smallest integer not greater than x

5. compute the approximate likelihood function (4.20)

6. find the *aml* estimates $\hat{\theta} = (\hat{b}, \hat{\mu}, \hat{\sigma})$ by maximizing the approximate likelihood function in step (5).

7. compute the estimation $\hat{\theta} - \theta$ error $\hat{\theta}_g - \theta$ and $\hat{\theta} - \theta$ for the estimates obtained both via Gaussian and MC approximation respectively.

Percentual relative errors for Gaussian and *aml* estimators are represented in Figures 4.5 for ma single run of the previous experiment corresponding to the sampling steps $\Delta = 70\sigma^2$.
The Gaussian and *aml* estimators are evaluated by averaging over the results of the MC replicates of the previous steps. In Table 4.1 the results of the Monte Carlo experiment are shown. The sample size is $n = 2000$. Relative bias (RBIAS) and relative root mean square error (RRMSE) are reported for the three sampling steps $\Delta = 7\sigma^2$, $\Delta = 70\sigma^2$, $\Delta = 700\sigma^2$, . For each sampling frequency, also the estimator $\hat{\hat{\sigma}}$ based on quadratic variation approximation (see formula 4.16) is evaluated. Moreover, error in the estimators of mean and standard deviation of the stationary distribution (*m.staz* and *std.staz*), obtained by plugging-in the parameter estimates in

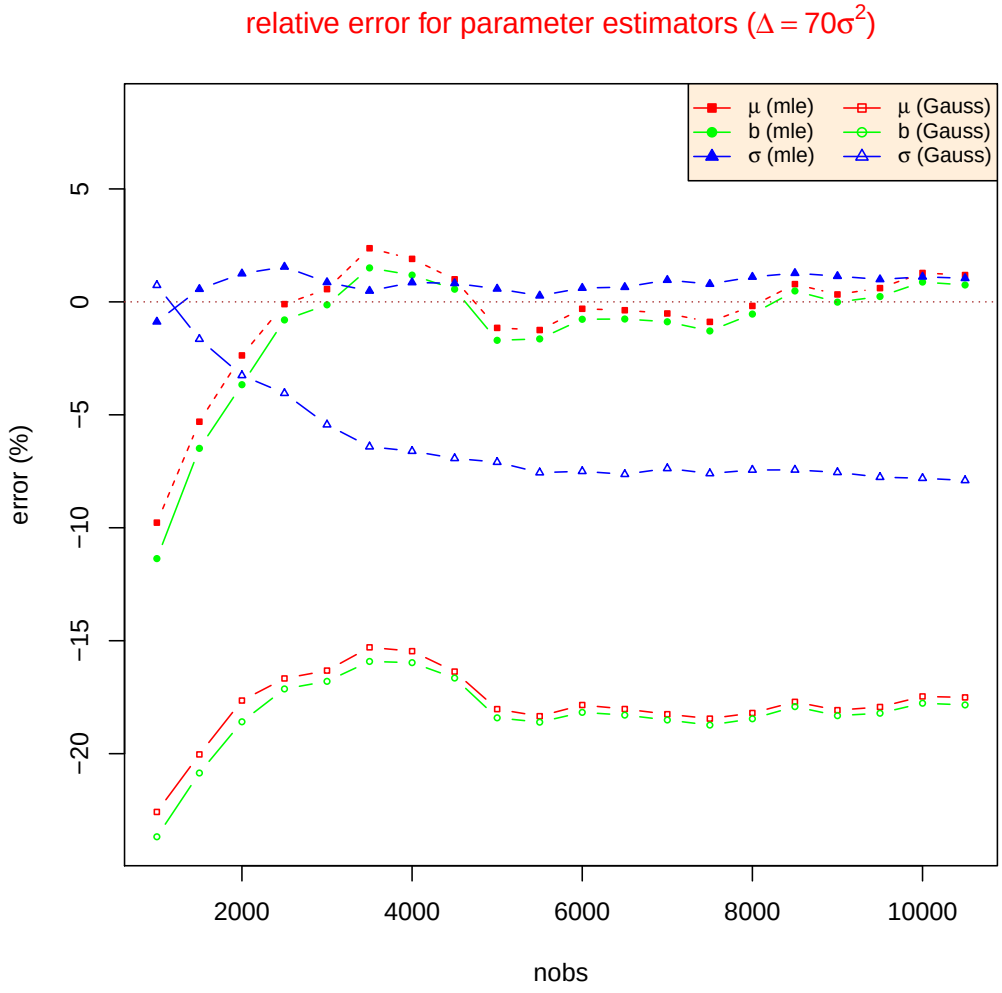

Figure 4.5: Relative error for the Gaussian and *aml* estimators.

the corresponding expressions is also reported for each of the two considered methods.

The results in Table 4.1 show that for small sampling step ($\Delta = 7\sigma^2$), performances of Gaussian estimator and *aml* estimator are quite similar. However, as the sampling frequency increases, the accuracy of the estimator based on the Gaussian approximation deteriorates becoming worse than the accuracy of the *aml* estimator. Moreover, comparison between the Gaussian estimator and the quadratic-variation based estimator of the volatility parameter σ shows that the latter is more accurate for the all the considered

Table 4.1: estimation error: Gaussian estimator (*Gauss*) vs. approximate mle (*amle*) for different sampling steps (Δ).

$\Delta = 7\sigma^2$		μ	σ	b	*m.staz*	*std.staz*
rbias	*Gauss*	0.01	0.04	0.01	0.00	0.06
	amle	0.02	0.00	0.03	0.00	0.00
	$\hat{\hat{\sigma}}$	-	0.00	-	-	-
rrmse	*Gauss*	0.13	0.05	0.14	0.00	0.08
	amle	0.14	0.02	0.15	0.00	0.07
	$\hat{\hat{\sigma}}$	-	0.00	-	-	-
$\Delta = 70\sigma^2$						
rbias	*Gauss*	-0.19	-0.16	-0.20	0.00	0.04
	amle	-0.02	0.00	-0.02	0.00	0.01
	$\hat{\hat{\sigma}}$	-	-0.08	-	-	-
rrmse	*Gauss*	0.20	0.16	0.20	0.00	0.06
	amle	0.06	0.02	0.06	0.00	0.03
	$\hat{\hat{\sigma}}$	-	0.09	-	-	-
$\Delta = 700\sigma^2$						
rbias	*Gauss*	-0.76	-0.61	-0.76	0.00	0.28
	amle	0.03	0.05	-0.03	0.00	0.04
	$\hat{\hat{\sigma}}$	-	-0.49	-	-	-
rrmse	*Gauss*	0.77	0.61	0.77	0.00	0.30
	amle	0.20	0.11	0.20	0.00	0.05
	$\hat{\hat{\sigma}}$	-	0.49	-	-	-

sampling steps and that its accuracy decreases as Δ increases. These results agree with the general theory illustrated in Chapter 3, where asymptotic theory for $\Delta \to 0$ is illustrated (see Section 3.3). Analysis of the last two columns of Table 4.1 shows that poor performance in estimating the model parameters does not reflect in a bad estimation of the first two moments of

the stationary distribution. In particular the estimator of the first moment is almost unbiased.

Table 4.2: estimation error for standardized parameters μ_s and b_s.

$\Delta = 7\sigma^2$		μ_s	b_s
rbias	*Gauss*	0.01	0.01
	amle	0.02	0.02
rrmse	*Gauss*	0.14	0.15
	amle	0.14	0.15
$\Delta = 70\sigma^2$			
rbias	*Gauss*	-0.05	-0.06
	amle	-0.02	-0.02
rrmse	*Gauss*	0.07	0.07
	amle	0.05	0.05
$\Delta = 700\sigma^2$			
rbias	*Gauss*	-0.15	-0.15
	amle	-0.06	-0.06
rrmse	*Gauss*	0.16	0.15
	amle	0.07	0.08

It is worth noting that in the third experiment, corresponding to the largest sampling step ($\Delta = 700\sigma^2$), the RRMSE of the approximate ML estimator is higher than in the case of the second experiment ($\Delta = 70\sigma^2$). This is probably due to the fact that, for very large sampling steps, the transition density function tends to coincide, as a function of its second spatial argument, with the density of the stationary ditribution. As we have remarked at the end of Section 1.2.3, the parameters of the latter are not separately identifiable. Specifically, in the present case, the stationary distribution density is invariant with respect to the transformations $\sigma^2 \to c\,\sigma^2, b \to c^2\,b, \mu \to c^2\,\mu;\ c > 0$. Hence, lack of identifiability for the stationary distribution reflects in low identifiability of the model parameters

for large sampling steps. The previous argument seems to be confirmed by the results reported in Tables 4.2, where, using the same experimental set up described above, the approximate ML and Gaussian estimators of the "standardized parameters" $\mu_s \doteq \mu/\sigma^2$, $b_s \doteq b/\sigma^2$ are evaluated. In fact, for the standardized parameters, the RRMSE of both estimators for $\Delta = 700\sigma^2$ is much smaller than for the original parameters, the approximate ML estimator being more accurate than the Gaussian estimator. As in the previous experiment, for the smallest sampling step ($\Delta = 7\sigma^2$), the performances of the two estimators are quite similar, while the Gaussian estimator is slightly less accurate than the competing estimator for $\Delta = 70\sigma^2$.

The previous argument suggests an alternative estimation strategy based on the equilibrium distribution. In fact, one could estimate the standardized parameters by maximizing the likelihood corresponding to the equilibrium density, based on a set of sufficiently "distant" observations, and, using an independent estimate of the volatility parameter σ (obtained, for instance, through Gaussian or quadratic variation approximation), to obtain estimates of the model parameters. In Table 4.3, results of such a strategy are reported. For each considered sampling step, the estimates of the model parameters based on the observations $u_1, \ldots, u_n$ are obtained as follows:

1. find the estimates $\hat{b}_s, \hat{\mu}_s$ of the standardized parameters b_s, μ_s by maximizing the likelihood function:

$$L_{st}(b_s, \mu_s) \doteq \mathcal{N}_0^{-n} \prod_{i=1}^{n} \frac{1}{u_i(K-u_i)^3} \left(\frac{u_i}{K-u_i}\right)^p \exp\left\{-2\mu_s \frac{u_i}{K-u_i}\right\}, \tag{4.21}$$

   where N is the normalization constant and $p = 2(b_s - \mu_s) - 1$ (see formula (1.26) in Section 1.2.3)

2. multiply the obtained estimates $\hat{b}_s, \hat{\mu}_s$ by the volatility estimate $\hat{\hat{\sigma}}$ obtained via the quadratic variation approximation.

Comparison of tables 4.1 and 4.3 shows that for large time interval between consecutive observations, the estimates of the standardized parameters are approximately as accurate as the *aml* estimates. However, poor accuracy of the volatility estimate causes bad estimates of the original parameters μ and b. For small sampling step ($\Delta = 7\sigma^2$), the quadratic variation based estimator of σ has low RRMSE, but due to the lack of independence between the observations, the estimates of the standardized parameters are biased.

Table 4.3: estimates based on the stationary distribution.

$\Delta = 7\sigma^2$	μ	σ	b	μ_s	b_s
rbias	-0.16	-0.00	-0.18	-0.15	-0.16
rrmse	0.21	0.02	0.22	0.20	0.21
$\Delta = 70\sigma^2$	μ	σ	b	μ_s	b_s
rbias	-0.22	-0.08	-0.22	-0.07	-0.08
rrmse	0.23	0.09	0.23	0.09	0.09
$\Delta = 700\sigma^2$	μ	σ	b	μ_s	b_s
rbias	-0.76	-0.49	-0.76	-0.08	-0.08
rrmse	0.76	0.49	0.76	0.08	0.08

In order to obtain approximate independence among the observations, one could subsampling the observed values of the process and use for estimation only values srongly separated in time. However, this would imply decrease of the sample size causing loss of efficiency.

The experimental results show that the relative errors for the estimates of the parameters μ and b are quite similar. In fact, it results that the estimation errors are strongly correlated (see Figure 4.5)

An interesting issue is the sensitivity of the results to the number of sub-intervals used for discretization in the algorithm. In Table 4.4 we report the parameter estimates corresponding to different discretization steps for a given realization of the SBM1 process with parameters as in the previous experiment. Again, three experiments are considered where the process is discretely observed ($n = 2000$ observations) at fixed time intervals between consecutive observations $\Delta = 7\sigma^2$, $\Delta = 70\sigma^2$, $\Delta = 700\sigma^2$ respectively. The different discretization steps correspond to different partitions of Δ in ($nint$) sub-intervals. The number of MC iterations for the likelihood approximation is fixed ($nsim = 20$). The values ($l.true$) that the approximate likelihood function takes on the true parameters is also reported for each discretization

Table 4.4: sensitivity of the estimates with respect to the discretization step. $\Delta = 7\sigma^2,\ n = 2000$ (Δ)

$\Delta = 7\sigma^2$	nint	μ	σ	b	$l.true$
	10	0.0433	0.1429	0.4081	7553.183
	20	0.0433	0.1428	0.4081	7552.022
	30	0.0432	0.1428	0.4077	7551.629
	40	0.0432	0.1428	0.4070	7551.415
	50	0.0432	0.1428	0.4070	7551.291
	60	0.0433	0.1428	0.4078	7551.239
	70	0.0432	0.1428	0.4075	7551.186
	80	0.0432	0.1428	0.4076	7551.144
	90	0.0432	0.1428	0.4076	7551.109
	100	0.0432	0.1428	0.4071	7551.072
$\Delta = 70\sigma^2$					
	10	0.0318	0.1432	0.3074	5641.321
	20	0.0317	0.1427	0.3067	5636.894
	30	0.0315	0.1425	0.3045	5634.522
	40	0.0317	0.1425	0.3057	5634.431
	50	0.0316	0.1424	0.3046	5633.557
	60	0.0316	0.1424	0.3052	5633.542
	70	0.0319	0.1426	0.3081	5635.008
	80	0.0315	0.1423	0.3042	5632.768
	90	0.0316	0.1424	0.3052	5633.128
	100	0.0314	0.1423	0.3034	5632.190
$\Delta = 700\sigma^2$					
	10	0.0680	0.2529	0.6527	5091.313
	20	0.0191	0.1158	0.1858	4966.004
	30	0.0203	0.1180	0.1973	4984.717
	50	0.0180	0.1112	0.1750	4939.406
	100	0.0177	0.1101	0.1723	4934.595
	160	0.0194	0.1143	0.1884	4968.043
	200	0.0188	0.1122	0.1821	4956.892
	500	0.0172	0.1082	0.1671	4918.850
	800	0.0175	0.1088	0.1698	4928.970
	1000	0.0174	0.1088	0.1688	4926.722

Table 4.5: sensitivity of the estimates of parameters μ_s, b_s with respect to the discretization step ($\Delta = 700$)

nint	μ_s	b_s
10	1.064	10.199
20	1.430	13.847
30	1.461	14.167
50	1.459	14.148
100	1.463	14.210
200	1.493	14.469
500	1.473	14.257
800	1.481	14.359
1000	1.472	14.291

step. From Table 4.4 it results that the estimates stabilize very soon for the small and the medium sampling steps, while a large number of discretization subintervals are needed in order the estimates to become stable for $\Delta = 700\sigma^2$. In particular, comparing the $\Delta = 70\sigma^2$ and $\Delta = 700\sigma^2$ experiments, where equal length sub-intervals correspond to *nint* values differing for a factor 10, it can be noticed that stabilization of the estimates is slower for the $\Delta = 700\sigma^2$ case. Major instability with respect to the discretization step for low frequency sampling case is probably due to the low identifiability of the parameters for large sampling step. This seems to be confirmed by the slightly higher stability in the estimates of the standardized parameters (see Table 4.5).

An other concern is the stabilization of the results with the number of MC iterations used for the likelihood approximation. The results of Monte Carlo experiments, not reported here, show that, for a fixed discretization step, the number of MC iterations needed to obtain stability of the estimates, increases as the sampling frequency becomes lower.
The analysis conducted on these numerical applications shows that in general, the approximate maximum likelihood estimators perform better than the competing estimators, specially when the sampling frequency is low. However, *aml* methodology requires much more computing time and memory resources and in some circustances, namely when the process is observed at high frequency, the benefits deriving from using *aml* methodology, instead of a much simpler method (like the one based on Gaussian approximation) do not justify so intensive computational efforts. Finally, it is worth mention the

optimization method used to maximize the approximate likelihood function. We recall that the latter is the product of the MC approximations of the transition densities evaluated on pairs of consecutive observations. Each approximating factor in the product is, in turn, a complicated function involving the MC average of discrete approximations of Brownian functional realizations (see step (3 of the algorithm in Section 4.2.2). Since the explicit form of the function to be maximize is quite complex, a maximization method that does not require computation of derivatives has been used. Specifically, *aml* estimates have been obtained through the methodology of bound optimization by quadratic approximation (*bobyqa*) as implemented in the R-package *minqa* (R Development Core Team, 2005).

All the procedures needed for the experimental applications are based on *ad hoc* R functions. In particular, functions for exact simulations and *aml* estimation of the SBM1 parameters have been implemented.

Concluding remarks

Modeling and forecasting the diffusion of innovations has received practical and academic interest since the 1960s. Pioneering works on this topics are those of Fourt and Woodlock (1960), Bass (1969), Mansfield (1961), dealing mainly with diffusion of new consumer durables, although the characteristics of the diffusion of new technologies or social phenomena are quite similar. One possible approach to the study of diffusion of innovations is to model the cumulative adoption, i.e., the total number of individuals that at a given time are adopters. Curve of cumulative adoption is typically S-shaped, and modeling efforts have been made in order to explain this behaviour. Models defined through logistic-type differential equations are particulary suitable to describe dynamic of innovation diffusions, in that they can account for the most important diffusion characteristics. In particular, the classical Bass model has been a reference model for many works in the marketing literature. In the Bass model, the main aspect of the diffusion, namely the interaction among different individuals of the population under investigation, is captured by the quadratic term in the differential equation governing the dynamic of the adoptions. This term is the product of the number of individuals who are already adopters by the number of individuals who have not adopted the innovation yet. Individual propension to adoption and disadoption are also taken into account through linear terms. In the early works, mainly in the marketing literature, diffusion has been often modeled via deterministic models. However, the need of introducing uncertainty in the analysis and forecasting diffusion of innovations has motivated significant reasearch in the area of statistical modeling. In this thesis, a stochastic extension of the deterministic Bass model has been studied. The extension has been obtained by means of the machinery of stochastic differential equations (SDE). In particular, the stochastic Bass model (SBM) is defined as a diffusion process which is solution of a SDE that directly extends the corresponding ordinary differential equation (ODE)in the classical theory. The stochastic term in the SDE has been defined in such a way that the solutions have (a.s.) a qualitative behaviour that allow to interpret them as continuous approximation

of number of adopters. In fact, exploiting some recent results (Shurz, 2007) on the qualitative behaviour of solutions of logystic-type SDEs in bounded domains, it is possible to define the stochastic Bass model in such a way that cumulative number of adoptions does not exceed a.s. the potential number K of adoptions. This property, referred to as "regularity" on the interval $[0, K]$, has been proved to hold more in general on time-dependent domains. More precisely, under the assumption of a dynamic of potential adopters represented by a (deterministic) non decreasing function $K(t)$, it has been shown that the solution $Y(t)$ of the resulting (non-homogeneous) Bass SDE belongs a.s. to the time-dependent interval $[0, K(t)]$ for each time t. This result could be useful in situations (not considered here) where interest is focused on potential adopters population dynamic. For instance, the function $K(\cdot)$ could contain unknown parameters to be estimated. The theoretical properties of the stochastic Bass model have been extensively studied in the central part of the thesis (Chapter 1). It has been shown that the process is ergodic with non degenerate invariant distribution when all the model parameters are positive. Furthermore, an explicit expression for the stationary distribution has been found involving modified Bessel functions of second kind. A supplemetary analysis has been carried out in the special case of stochastic Bass model without self-innovation (SBM1). In this case, it has been proved that ergodicity holds only under certain restrictions among the model parameters. Specifically, if the noise parameter σ is large enough (with respect to the difference between imitation and disadoption parameter), an invariant distribution does not exist. A detailed analysis has been conducted on the stochastic stability of trivial solutions of the SDE corresponding to the SBM1, when the model parameters are such that the process is not ergodic. Using a suitably defined Lyapunov function, it has been shown that the trivial solution $Y_t \equiv 0$ can be (asymptotically) stable in probability even in situations where the corresponding deterministic solution ($\sigma = 0$) is not, i.e., when the imitation coefficient b is greater than the disadoption coefficient μ. In other words, similarly to what happens for other logistic models used, for instance, for the study of population dynamics, noise can determine extintion.

Introduction of uncertainty in models for the diffusion of innovations through stochastic differential equations poses serious problems that have to be faced in order this modeling to be of practical interest. A central problem concerns the estimate of the model parameters. This is related to the general issue of statistical inference on diffusion processes discretely observed, a topic that has received growing interest during the last three decades. Research on this topic has moved along different directions. Among various

approaches, we mention numerical solution of the Fokker-Planck equation, estimating functions, indirect inference, and likelihood-based methods. A review of the main approaches to the statistical inference for diffusion processes is contained in Chapter 3. Major emphasis is posed on methods based on likelihood function. If the stochastic process were continuously oberved, inference could rely on the *continuous time likelihood*, provided that the diffusion term is known. Thus, some approaches exploit the relation between diffusion term in SDE and quadratic variation of the solutions, to estimate parameters in the diffusion term, and then, treating these parameters as known, approximate continuous time likelihood with some discretization procedure based on the observed values. However, as remarked in Chapter 3, this approach can determine serious bias if the process is not observed at sufficiently high frequency. In fact, Florens-Zmirou (1989) has shown that for fixed time intervals between consecutive observations, the resulting estimator is not consistent as the number of observations becomes infinite.

In the thesis, another approach has been adopted. In fact, inferencial procedures have been considered that directly rely on the discrete nature of the observation process. More precisely, estimation is based on the *discrete time likelihood*. By Markov property, the latter is simply the product of the values that the transition density function takes on pairs of consecutive times/observations. The problem with this approach is that for the stochastic Bass model, as for most diffusion processes implicitely defined through SDEs, the analytic form of the transition density function is not available. Thus, estimation procedures must rely on numerical approximations. In this thesis, two types of numerical related schemes have been studied and implemented. The first one, based on the recent work of Beskos, Papaspilioupolos and Roberts (2006), concerns "exact" simulation of diffusion process. The methodology, based on the *retrospective sampling* has been described in general in Chapter 2. The specific adaptation to the stochastic Bass model has been presented in Chapter 4. The second class of numerical methods includes methods for MC approximation of the transition density function. This class of methods has been illustrated in general in Chapter 3.

In Chapter 4 parameter estimation via MC approximation of the transition density function has been analyzed for the SBM1. The approach adopted is based on the relation between the transition density of the (suitably transformed) process and the expectation of a particular Brownian functional. Evaluation of the methodology has been performed by comparing the resulting estimator with the classical estimator based on Gaussian approximation of the transition density. As expected, it results that if the process

is observed at high frequency, the accuracy of the estimators based on the Gaussian approximation is similar to the one of the estimators based on the MC approximation. Thus, in these circustamces the Gaussian estimator is prefearable, because of its simplicity. However, as the sampling step increases, the accuracy of the Gaussian estimator rapidly deteriorates becoming much worse than the competing estimator. These experimental results agree with the general asymptotic results mentioned in Chapter 3. On the other hand, when the process is sampled at very low frequency, also the MC estimator of the model parameter provides poor performaces. One possible explanation for this behaviour is that, as the time interval between consecutive observations increases, these observations tend to become independent one of each other. In these situations, the observed values of the process can be (approximately) viewed as a sample from the invariant distribution of the process. Since density function of this distribution dependens on the volatility parameter σ only through the ratios involving the "deterministic" parameters $a/\sigma^2, b/\sigma^2\mu/\sigma^2$, the parameters are not separately identifiable. This is confirmed by the fact that, in spite of the lack of identifiability, the estimators of the first two moments of the stationary distribution seem to be consistent.

The estimation strategy adopted in this thesis relies on the Girsanov theorem. In this context, it is used to express the transition density of the target diffusion process in terms of the Radon Nikodym derivative of the its law with respect to the Wiener measure. The resulting expression contains an integral Brownian functional, which is approximated with a discrete usm whose terms are randomly drawn from a suitable finite dimensional law of the Brownian Bridge. Discretization may imply bias, and its impact should be evaluated in specific cases. In certain circustamces, basically the same that must hold for the applicability of the exact algorithm, an umbiased estimator can also be used in the spirit of the SAM method (see Section 3.5.5).

Another approach is based on explicit simulation of diffusion paths. In this framework, two classes of methods can be distinguished depending on whether unconditional simulation (Section 3.5.2) or diffusion bridge simulation (Section 3.5.3) is required. Methods belonging to the latter class are more efficient because they are based on a data augmentation strategy that explicitly take into account observed values. However, simulation of paths conditioned to the ending points of an interval is more difficult than unconditional simulation. In fact, unconditional simulation, differently from conditional simulation, can use simple Euler-type schemes. An elegant and efficient method for diffusion bridge simulation is the conditioned exact algo-

rithm (see Section 2.2.6). It is an acceptance-rejection scheme that provides sample paths with given starting and ending points, from the true target bridge measure. However, differenltly from unconditional simulation methods, the time interval on which the diffusion bridge is defined, cannot be broken in sub-intervals in order to incerase the accepyance probability. Thus, the applicability of the algorithm within inferential approaches based on data augmentation schemes depends on the time distance between consecutive observations.

Extension of the methodology to multivariate diffusions is highly non trivial. In fact, two important limitations make the methodology dificult to apply. First, the tranformation needed to "standardize" the stochastic process, i.e., to reduce it to a unit volatility process, requires that the rows of the inverse of the volatility matrix are conservative fields (see formula (2.36) of Chapter 2). Second, in order the Brownian integral in the Girsanov formula to be eliminated, the drift function in the trasformed SDE must be of gradient type. These two requirements seriously limit the class of multivariate diffusions for which exact simulation (and related inferential methods) is applicable.

In this thesis, inference on diffusion processes has been considered in situations where he process is observed without errors. However, filtering of discretely observed diffusion processes is also of great interst. A recent paper on this topic, involving exact simulation methodology is that of Fearnhead et al. (2008). In this paper, the authors use retrospective sampling in combination with the particle filter methodology (Pitt and Sheppard, 1999; Doucet et al., 2001; Chopin, 2004) to approximate the filtering density of interest. A variety of observation schemes are considered including partial observation of the components of a diffusion process and arrival times of a Cox process. In Fearnhead et al. (2008), the diffusion parameters, as well as the parameters of the observation proces are assumed to be known. An interesting topic is the extension of the approach to include filtering problems with unknown parameters.

Appendix A

Rejection Sampling Algorithm

Proposition 3 *Let $(S, \mathcal{A})$ a probability space and μ, ν, two measures on S with the property that there exists $\epsilon > 0$ such that*

$$f \doteq \epsilon \frac{d\mu}{d\nu} \leq 1, \quad \nu \quad a.s.$$

Let $(Y_n, I_n)_{n\geq 1}$ be a sequence of i.i.d. random elements taking values in $S \times \{0, 1\}$, such that $Y_1 \sim \nu$, and:

$$P\{I_1 = 1 | Y_1 = y\} = f(y), \ \ \forall y \in S.$$

Define $\tau = \min\{i \geq 1 | I_i = 1\}$. Then $P\{Y_\tau \in dy\} = \mu(dy)$.

Proof.

$$\begin{aligned} P\{I_i = 1\} &= \int_S P\{I_1 = 1 | Y_1 = y\} \nu(dy) = \int_S f(y)\nu(dy) \\ &= \int_S \epsilon \frac{d\mu}{d\nu} \nu(dy) = \epsilon \int_S \mu(dy) = \epsilon. \end{aligned}$$

For $\mathcal{F} \in \mathcal{A}$, we have:

$$
\begin{aligned}
P\left\{Y_\tau \in \mathcal{F}\right\} &= P\left\{Y_\tau \in \mathcal{F}, I_1 = 1\right\} + P\left\{Y_\tau \in \mathcal{F}, I_1 = 0\right\} \\
&= P\left\{Y_\tau \in \mathcal{F}, I_1 = 1\right\} + P\left\{Y_\tau \in \mathcal{F} | I_1 = 0\right\} P\left\{I_1 = 0\right\} \\
&= P\left\{Y_\tau \in \mathcal{F}, I_1 = 1\right\} + P\left\{Y_\tau \in \mathcal{F}\right\} P\left\{I_1 = 0\right\},
\end{aligned}
$$

where the last equality follows from the independence of the random elements $(Y_n, I_n)_{n\geq 1}$ for different values of i and the fact that if $I_i = 0$, then $\tau > 1$. Since $I_1 = 1$ implies $\tau = 1$, it follows that the event $\{Y_\tau \in \mathcal{F}, I_1 = 1\}$ is the same event as $\{Y_1 \in \mathcal{F}, I_1 = 1\}$. Thus:

$$
\begin{aligned}
P\left\{Y_1 \in \mathcal{F}, I_1 = 1\right\} &= \int_{\mathcal{F}} P\left\{I_1 = 1 | Y_1 = y\right\} \nu(dy) \\
&= \int_{\mathcal{F}} f(y)\nu(y) = \epsilon\mu(\mathcal{F}).
\end{aligned}
$$

Finally we obtain:

$$
\begin{aligned}
P\left\{Y_\tau \in \mathcal{F}\right\} &= \epsilon\mu(\mathcal{F}) + P\left\{Y_\tau \in \mathcal{F}\right\} P\left\{I_1 = 0\right\} \\
&= \epsilon\mu(\mathcal{F}) + P\left\{Y_\tau \in \mathcal{F}\right\} (1 - \epsilon).
\end{aligned}
$$

It follows that $P\left\{Y_\tau \in \mathcal{F}\right\} = \mu(\mathcal{F})$ $\diamondsuit$

Note that the previous scheme does not assume any order in the simulation of the "decision variable" (I) and the target variable (Y). This circustamce allows to implement a a acceptance-rejection scheme for simulation of diffusion process based only on a finite number of instances of sample paths.

Appendix B

The Brownian Bridge

Let W be a standard Brownian motion on the filtered space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$. The Brownian Bridge $BB_{(x,S)\to(y,T)}$, between (x, S) and (y, T) is the stochastic process defined as the process W conditioned on taking values x, y at times S, T respectively. Without loss of generality we will assume $S = 0$ and denote the corresponding Brownian Bridge as $BB^{T,x,y}$. The Brownian Bridge $BB^{1,0,0}$. is called *standard Brownian Bridge* and is shortly denoted by BB.

Properties.

- *Gaussianity.*

  The Brownian Bridge $BB^{T,x,y}$ is a Gaussian process with expected value and covariance:

$$\begin{aligned} E\left[BB_t^{T,x,y}\right] &= x + \frac{t}{T}(y - x); \\ Cov\left[BB_s^{T,x,y},\, BB_t^{T,x,y}\right] &= s \wedge t - \frac{st}{T} \end{aligned}$$

- *relocation invariance property.*

  If $BB_t^{T,0,0}$ is a Brownian Bridge between $(0, 0)$ and $(0, T)$, then

$$BB_t^{T,x,y} = BB_t^{T,0,0} + \left(1 - \frac{s}{t}\right) x + \frac{s}{t} y$$

  is a Brownian Bridge between $(x, 0)$ and (y, T).

- *characterization via SDE.*

  The Brownian Bridge $BB_t^{T,x,y}$ can be defined as the solution of the linear SDE:

  $$dY_t = \frac{y-T}{T-t}dt + dW_t$$

  and thus can be represented as:

  $$BB_t^{T,x,y} = \left(1 - \frac{s}{t}\right)x + y\frac{t}{T} + (T-t)\int_0^t \frac{dW_s}{T-s}$$

- *relation with standard Brownian motion.*

  The Brownian Bridge $BB_s^{T,0,0}$ can be obtained as trasformation of the standard Brownian motion W:

  $$BB_s^{T,0,0} = W_s - \frac{s}{T}W_T.$$

Appendix C

Stochastic stability

Consider the SDE

$$dY_t = b(Y_t)dt + \sigma(Y_t)dW_t, \tag{C.1}$$

and assume that $b(0) = \sigma(0) = 0$, so that $Y_t \equiv 0$ is a solution of the SDE. In order to make notation simpler, we will use the notation P^y to denote the probability measure induced on $\mathcal{C}(0,T)$ by the solution of the SDE (C.1) with initial condition $Y(0) = y$.

Definition 1 *The solution $Y_t \equiv 0$ is said to be **stochastically stable**, or **stable in probability**, if*

$$\lim_{y\to 0} P^y \left\{ \sup_{t\geq 0} |Y_t| > \epsilon \right\} = 0 \tag{C.2}$$

holds for any $\epsilon > 0$.

The definition (1) means that we can make arbitrarily small the probability of escape from a (arbirarily) small right neighborood of the origin, provided that the initial condition y is close enough to the equilibrium condition $y = 0$. In other words, for given ϵ, the probability $p^\epsilon(y)$ of escape from $(0,\epsilon)$ for the solution with initial condition y, is a continuous function at $y = 0$.
The following definition introduces a stronger notion of stability.

Definition 2 *The solution $Y_t \equiv 0$ is said to be* ***stochastically asymptotically stable*** *if it is stochastically stable and if*

$$\lim_{y\to 0} P^y \left\{ \lim_{t\to\infty} Y_t = 0 \right\} = 1. \tag{C.3}$$

If $P^y \{\lim_{t\to\infty} Y_t = 0\} = 1$, for *any* y, then we say that the zero solution is *globally asymptotically stable* in probability.

As for the case of ODEs, a useful tool for studying stochastic stability is the *Lyapunov function* that we define below.

Let U be an open set in $\mathbb{R}^n$ A function V: $U \to \mathbb{R}$ is *proper* if it satisfies:

$$f(|x|) \leq V(x) \leq g(|x|)$$

for some strictly increasing functions f and g which satisfy $f(0) = g(0) = 0$.

A Lyapunov function is a proper continuous function which is C^2 on $U \setminus \{0\}$.

In the deterministic context, given the autonomous system $\dot{x} = F(x)$, the stability (asymptotic stability) of an equilibrium point x^* ($F(x^*) = 0$), can be established if one can find a Lyapunov function V such that $\dot{V}(x) \leq 0$ ($\dot{V}(x) < 0$) in a neighborood U of zero (here the dot notation denotes derivative with respect to time). The Lyapunov criterion for stability has the following generalization to the stochastic case Has'minski (1980).

Let $\mathcal{L}$ be the generator of the SDE (C.1), i.e.:

$$\mathcal{L} = b(y)\frac{\partial}{\partial y} + \frac{1}{2}\sigma^2(y)\frac{\partial^2}{\partial y^2},$$

and assume that a Lyapunov function V exists such that

$$\mathcal{L}V(y) \leq 0 \tag{C.4}$$

for $y \in U \setminus \{0\}$; then the equilibrium solution $Y_t \equiv 0$ is stochastically stable. If inequality (C.4) is strict, asymptotic stochastic stability holds.

Bibliography

Ait-Sahalia, Y. (2002), 'Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach', *Econometrica* **70**, 223–262.

Ait-Sahalia, Y. (2008), 'Closed form likelihood expansions for multivariate diffusions', *Annals of Statistic* **36**(2), 906–937.

Asmussen, S., Glynn, P. and Pitman, J. (1995), 'Discretization error in simulation of one-dimensional reflecting Brownian motion.', *Ann. Appl.Probab.* **5**(4), 875–896.

Bass, F. M. (1969), 'A new product growth model for consumer durables', *Management Sciences* **15**, 215–227.

Beskos, A., Papaspiliopoulous, O. and Roberts, G. O. (2006), 'Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes', *J.R. Statist. Soc. B* **68**(2), 1–29.

Beskos, A., Papaspiliopoulous, O. and Roberts, G. O. (2009), 'Monte Carlo maximum likelihood estimation for discretely observed diffusion proceses', *The Annals of Statistics* **37**, 223–245.

Beskos, A., Papaspilioupolos, O. and Roberts, G. O. (2006), 'Retrospective Exact Simulation of Diffusion Sample Path with Applications', *Bernoulli* **12**, 1077–1098.

Billingsley, P. (1961), 'The Lindberg-Levy theorem for martingales', *Proc. Amer. Math. Soc.* **12**, 788–792.

Chan, K. S. and Ledolter, J. (1995), 'Monte Carlo EM estimation for time series models involving counts', *J. Am, Statist. Ass.* **90**(1), 242–252.

Chopin, N. (2004), 'Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference', *Annals of Statistic* **32**, 2385–2411.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), 'Maximum likelihood from incomplete data via the em algorithm', *J.R. Statist. Soc. B* **39**, 1–38.

Devroye, L. (1986), *Non uniform random variate generation*, Springer-Verlag, New York.

Doucet, A., de Freitas, N. and Gordon, N. (2001), An introduction to sequential Monte Carlo methods, *in* 'Sequential Monte Carlo methods in Practice', Springer, New York, pp. 1–14.

D.P.Giesy (1976), Strong laws of large numbers for independent sequences of banach space-valued random variables, *in* B. Springer, ed., 'Probability in Banach Spaces (Proc. First Internat. Conf., Oberwolfach,1975). Lectures Notes in Math.', Vol. 526, pp. 89–99.

Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. O. (2008), 'Particle filters for partially observed diffusions', *J.R. Statist. Soc. B* **70**(4), 755–777.

Florens, D. (1999), 'Estimation of diffusion coefficient from crossings', *Statistical Inference for Stochastic Processes* **1**, 175–195.

Florens-Zmirou, D. (1989), 'Approximate discrete-time schemes for statistics of diffusion processes', *Statistics* **20**, 547–557.

Fort, G. and Moulines, E. (2003), 'Convergence of the Monte Carlo expectation maximization for curved exponential families', *Ann. Statist.* **31**, 1220–1259.

Fourt, L. A. and Woodlock, J. W. (1960), 'Early prediction of early success of new grocery products', *Journal of Marketing* **25**, 31–38.

Gourieroux, C., Monfort, A. and Renault, E. (1993), 'Indirect Inference', *J. Appl. Econometrics* **8**, 85–118.

Gutierrez, R., Nafidi, A. and Sanchez, R. G. (2005), 'Forecasting natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model', *Applied Energy* **80**, 115–124.

Has'minski, R. (1980), *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff.

Heyde, C. C. (1997), *Quasi-Likelihood and its applications. A General Approach to Optimal Parameter Estimation*, Springer-Verlag, New York.

Kanniainen, J., Makinen, S. J., Piche, R. and Chakrabarti, A. (2011), 'Forecasting the Diffusion of Innovation: A Stochastic Bass Model With Log-Normal and Mean-Reverting Error Process', *IEEE transaction om engineering management* **58**(2), 228–249.

Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.

Kessler, M. (1997), 'Estimation of an ergodic diffusion from discrete observations', *Scand. J. of Statist.* **24**, 211–229.

Kessler, M. (2000), 'Simple and explicit estimating functions for a discretely observed diffusion process', *Scand. J. of Statist.* **27**, 65–82.

Kloeden, B. and Platen, E. (1995), *Numerical Solution of Stochastic Differential Equations*, Springer, New York.

Lipster, R. and Shiryayev, A. (1977), *Statistics of Random Processes, volume 1*, Springer-Verlag, New York.

Mahajan, V. and Wind, Y. (1986), *Innovation Diffusion Models of New Products Acceptance*, Ballinger Pub. Co., Cambridge, Massachusetts.

Mansfield, E. (1961), 'Technical change and the rate of imitation', *Econometrica* **29**, 741–766.

Maruyama, G. (1955), 'Continuous Markov Processes and Stochastic Equations', *Rend. Circ. Mat. Palermo* **4**, 48–90.

Moldovan, S. and J.Goldenberg (2003), 'Cellular automata modeling of resistance to innovations: Effects and solutions', *Technological Forecasting and Social Change* **71**, 425–442.

Niu, S. C. (2002), 'A stochastic formulation of the Bass model of new-product diffusion', *Mathematical Problems in Engineering* **8**, 249–263.

Oksendal, B. K. (1998), *Stochastic Differential Equations: An Introduction with Applications*, Springer-Verlag, Berlin.

Papaspiliopoulos, O. and Roberts, G. O. (2009), Importance sampling techniques for estimation of diffusion models, Technical Report 28, CRISM - Centre for Research in Statistical Methodology.
URL: *www.warwick.ac.uk/go/crism*

Parker, P. M. (1994), 'Aggregate diffusion forecasting models in marketing: a critical review', *Int. J. Forecasting* **10**, 353–380.

Pascucci, A. (2008), *Calcolo Stocastico per la Finanza*, Springer Verlag Italia, Milano.

Pedersen, A. R. (1995), 'Consistency and asymptotic normality of an approximate likelihood estimator for discretely observed diffusion processes', *Bernoulli* **1**, 257–279.

Pitt, M. K. and Sheppard, N. (1999), 'Filtering via simulation: auxiliary particle filters', *J. Am. Statist.Ass.* **94**, 590–599.

Polson, N. and Roberts, G. (1994), 'Bayes factors for discrete observations for diffusion proceses', *Biometrika* **81**, 11–26.

Poulsen, R. (1999), Approximate maximum likelihood estimation of discretely observed diffusion processes, Technical Report 29, Centre for Analytical Finence, Aarhus.

Prakasa-Rao, B. L. S. (1983), 'Asymptotic theory for non-linear least square estimator for diffusion processes', *Math Operationsforsch. Statist. Ser. Stat* **14**, 195–209.

R Development Core Team (2005), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: *http://www.R-project.org*

R.Guseo and M.Guidolin (2010), 'Cellular Automata with Network Incubation in Information Technology Diffusion', *Physica A: Statistical Mechanics and its Applications* **389**(12), 2422–2433.

Risken, H. (1996), *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer Series in Synergetics, 2nd edition, Springer.

Roberts, G. O. and Stramer, O. (2001), 'On inference for partially observed diffusion models using the Metropolis-Hastings algorithm', *Biometrika* **88**, 603–621.

Rogers, E. (1995), *Diffusion of Innovations*, The Free Press, New York.

Rogers, L. and D.Williams (2000), *Diffusions, Markov Processes and Martingales. Vol 2: Ito Calculus*, Cambridge University Press, Cambridge.

Sherman, R., Y.K., H. and Dalal, S. (1999), 'Condition for convergence of Monte Carlo sequences with an application to product diffusion modeling', *Econ. J.* **2**, 248–267.

Shurz, H. (2007), 'Modeling, analysis and discretization of stochastic logistic equations', *International Journal of Numerical Analysis and Modeling* **4**, 178–197.

Shuster, J. (1968), 'On the Inverse Gaussian Distribution Function', *Journal of the American Statistical Association* **63**(324), 1514–1516.

Skiadas, C. H. and Giovanis, A. (1997), 'A stochastic Bass innovation diffusion model for studying the growth of electricity consunption in Greece', *Applied Stochastic Models and Data Analysis* **13**, 95–101.

Sorensen, M. (1999*a*), 'On asymptotics of estimating functions', *Braz. J. Probab. Statist.* **13**, 111–136.

Sorensen, M. (1999*b*), 'Prediction-based estimating functions', *Econometrics J.* **3**, 123–147.

Tanner, M. and Wong, W. H. (1987), 'The calculation of posterior distributions by data augmentation (with discussion)', *J.Am.Statist. Assoc.* **82**, 528–550.

Wei, G. and Tanner, M. (1990), 'A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms', *J. Am. Statist.Ass.* **85**, 699–704.

Yoshida, N. (1992), 'Estimation for diffusion processes from discrete observations', *J.Multivar. Anal.* **41**, 220–242.