

Exact Simulation of Diffusions and new Inference Methods for Discrete Time Data;

also included, the One-Shot CFTP Algorithm

Alexandros Beskos

Submitted for the degree of Doctor of Philosophy

at Lancaster University,

March 2006.

The Alternative Format

This is a thesis developed under the “alternative format”. It is comprised mostly of six articles, three of which have been submitted and the rest have been accepted for publication by international journals of scientific research. The articles are presented in the thesis in the time order that they came out, starting with the earliest one. For convenience, they are labelled as Article 1, Article 2, etc.

Article 1 and Article 2 were developed jointly with my supervisor, Professor Gareth O. Roberts, and are to be published at the journal of “Methodology and Computing in Applied Probability” and the journal of “Annals of Applied Probability” respectively. Article 4 has been accepted for publication as a discussion paper at the “Journal of the Royal Statistical Society, series B (Statistical Methodology)” and was developed in collaboration with my supervisor, Dr. Omiros Papaspiliopoulos and Dr. Paul Fearnhead. The other three articles still awaiting the judgement of the referees, were written jointly with my supervisor and Dr. Omiros Papaspiliopoulos.

In addition to the six articles, I have included in the thesis three short reviews, one for each of the major subjects that our articles are dealing with. The presentation follows an intuitive approach without emphasis on technical details. The objective has been to offer the reader a briefing on the general scientific area that our own methods are placed into, describe the methodological problems that our research dealt with and present methods suggested by other researchers. The reviews will also examine deficiencies of alternative methods and try to emphasize

the importance of our own contributions.

Abstract

Our research began by studying the Perfect Simulation idea developed by Propp and Wilson. In Article 1 we present the “One-Shot CFTP” framework, a general Perfect Simulation algorithm that constructs a Markov chain with a given target density as its equilibrium density and then carries out Perfect Simulation on this chain. The method is based on the Read-Once CFTP idea of Wilson and the Gibbs sampler. It synthesises coupling methods having appeared in the literature but also involves a new coupling technique we have developed and which we call the “Monotone Independence Coupler”. We apply the One-Shot CFTP idea to a family of truncated gaussian densities and demonstrate its computational efficiency even in high dimensions.

We then tried to apply Perfect Simulation ideas in the context of diffusion processes. Yet, the algorithm we came up with, termed the “Exact Algorithm” (EA), is based simply on rejection sampling. EA, when applicable, returns complete paths of a diffusion process through an accept/reject mechanism with Brownian path candidates. In Article 2 we present its initial version. We later managed to simplify the algorithm and to weaken considerably its conditions. The improved algorithms appear as EA1 and EA2 in Article 3 and EA3 in Article 6. In Article 6 we also present an algorithm for the exact simulation of Brownian motion hitting times for arbitrary one-sided smooth boundaries.

We exploit the properties of EA to carry out parametric inference for discretely observed diffusion processes. We present methods for the unbiased estimation of

the likelihood function, a Markov chain Monte Carlo algorithm and an EM approach. All methods build on the exact simulation of the diffusion paths provided by EA and avoid systematic errors that characterise alternative methods. They also appear to be computationally efficient. In Article 4 we introduce all our inference methods. In Article 5 we expand on the unbiased estimators of the likelihood surface and present the theory that justifies the properties of the related parameter estimators.

Acknowledgements

I am deeply grateful to my supervisor, Professor Gareth Roberts, for his guidance and friendship. It has been a pleasure and fortune working with him. I thank Dr. Omiros Papaspiliopoulos for his help, advice and support. The team the three of us formed has been extremely co-operative and productive. I am thus indebted to Professor Petros Dellaportas from my undergraduate department in Greece who advised me to move to Lancaster to work with Gareth and Omiros.

I also thank Bruno Casella, a late arrival to our team, Theodore Kypraios and Kostas Kalogeropoulos for many fruitful discussions.

I am grateful to my parents for their constant support and patience. Though not from an academic background, they showed great appreciation to my effort.

I thank the Greek State Scholarship's Foundation (in Greek, IKY) for funding my studies at Lancaster University.

Since this thesis is comprised mostly of articles already submitted for publication to scientific journals, I thank all the unknown referees for comments and corrections which have been incorporated in this thesis.

Declaration

I hereby declare that this thesis is my own work, except otherwise stated, and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Articles 1 and 2 were written jointly with my supervisor, Professor Gareth Roberts, and are now to appear in “Methodology and Computing in Applied Probability” and “Annals of Applied Probability” respectively. The computer code and the simulation results in both articles are my own work. Article 4 was written jointly with my supervisor and also Dr. Omiros Papaspiliopoulos and Dr. Paul Fearnhead. The computer code in the article is my own work. The article is to appear as a discussion paper in the “Journal of the Royal Statistical Society, series B (Statistical Methodology)”. Articles 3, 5 and 6 were written jointly with my supervisor and Dr. Omiros Papaspiliopoulos and have been submitted for publication to international journals of scientific research. The computer code included in these articles is my own work.

Alexandros Beskos

Professor Gareth Roberts

Contents

I	Introduction	1
A	The CFTP Algorithm	2
A.1	Introduction	2
A.2	The CFTP protocol	4
A.2.1	CFTP and stochastic monotonicity	8
A.2.2	Applications of CFTP	10
A.3	Reversing the time: Read-Once CFTP	10
A.4	Dominated CFTP	13
A.5	Coupling methods	17
A.5.1	Multigamma coupling	18
A.5.2	Rejection coupling	19
A.5.3	Catalytic coupling	20
A.5.4	Layered multishift coupling	21
A.6	Some general comments	23
A.7	Our contribution	24
B	Approximate Simulation Methods for SDEs	26
B.1	Introduction	26
B.2	Strong Taylor approximations	28
B.2.1	The general Taylor scheme	30
B.3	Other families of strong schemes	33

B.3.1	Derivative free schemes	33
B.3.2	Implicit strong schemes	34
B.4	The weak convergence criterion	36
B.5	Some general comments	38
B.6	Our contribution	39
C	Parametric Inference for Discretely Observed Diffusions	40
C.1	Introduction	40
C.1.1	Preliminary notation and definitions	42
C.2	Likelihood approximations	43
C.2.1	A closed form approximation	43
C.2.2	Simulated maximum likelihood	46
C.2.3	Comments	48
C.3	Estimating functions	49
C.3.1	Martingale estimating functions	50
C.3.2	Other families of EFs	52
C.3.3	Comments	53
C.4	Bayesian-MCMC Methods	54
C.4.1	High frequency augmentation (HFA)	54
C.4.2	Missing path transformation	56
C.4.3	Comments	57
C.5	Auxiliary model methods	57
C.5.1	Indirect Inference	59
C.5.2	Efficient Method of Moments (EMM)	60
C.5.3	Comments	61
C.6	Some general comments	62
C.7	Our contribution	63

II Articles 64

1	One-Shot CFTP; Application to a Class of Truncated Gaussian Densities	65
1.1	Introduction	65
1.2	One-Shot CFTP	69
1.3	The Monotone Independence Coupler	73
1.3.1	MIC in the context of TN s	77
1.4	The case-specific algorithm	78
1.5	The pure algorithm	85
1.6	Extensions	92
1.6.1	Extensions within the gaussian family	92
1.6.2	Further directions for extensions	94
1.7	Simulation results	98
1.7.1	Running the algorithms	98
1.7.2	Efficiency evaluation	100
1.8	Conclusions	102
2	Exact Simulation of Diffusions	108
2.1	Introduction	108
2.2	A general rejection sampling algorithm	110
2.3	The Exact Algorithm	111
2.3.1	Rejection sampling for SDEs	112
2.3.2	Constructing the Exact Algorithm	115
2.3.3	A factorisation for $\mathbb{Q}$	121
2.4	Efficiency of the Exact Algorithm	122
2.5	Applying the Exact Algorithm	125
2.6	Exact sampling of extremes and hitting times.	128
2.6.1	An application	132

2.7	Extensions and conclusions	133
3	Retrospective Exact Simulation of Diffusion Sample Paths with Applications	140
3.1	Introduction	140
3.2	Retrospective rejection sampling	143
3.2.1	The case when ϕ is bounded	146
3.3	The case when $\limsup_{u \rightarrow \infty} \phi(u) < \infty$	147
3.3.1	Decomposing the Brownian path at its minimum	148
3.3.2	The extension of the Exact Algorithm	151
3.4	The efficiency of the Exact Algorithm	152
3.4.1	User-impatience bias	154
3.5	Application: the logistic growth model	155
3.6	Applications of the Exact Algorithm	157
3.6.1	Monte Carlo maximum likelihood for discretely observed diffusions	157
3.6.2	Exact simulation of conditioned diffusions	164
3.7	Conclusions	165
4	Exact and Computationally Efficient Likelihood-Based Estimation for Discretely Observed Diffusion Processes	171
4.1	Introduction	171
4.2	Retrospective exact sampling of diffusions	176
4.2.1	Exact Algorithm 1 (EA1)	181
4.2.2	Exact Algorithm 2 (EA2)	182
4.2.3	Unconditional EA and path reconstruction	184
4.3	Unbiased transition density estimators – preliminaries	184
4.4	The bridge method	186
4.5	The acceptance method (AM)	187

4.5.1	The simultaneous acceptance method (SAM)	188
4.6	The Poisson estimator	191
4.6.1	Inference for the CIR model	193
4.7	Comparison of transition density estimators	194
4.8	A Monte Carlo EM (MCEM) approach	197
4.8.1	MCEM for diffusions with known diffusion coefficient	198
4.8.2	The general case	199
4.9	Hierarchical simulation models and inference using MCMC	201
4.10	Conclusions	206
5	Monte Carlo Maximum Likelihood Estimation for Discretely Observed Diffusion Processes	215
5.1	Introduction	215
5.2	Notation and statement of the main result	219
5.3	Uniform convergence of the Monte Carlo estimator	222
5.4	Numerical illustration	225
5.5	Proof of Theorem 5.1	226
5.6	Discussion	232
6	A Factorisation of Diffusion Measure and Finite Sample Path Constructions	235
6.1	Introduction	235
6.2	The Wiener-Poisson factorisation of $\mathbb{Q}$	237
6.2.1	Exact Algorithm 1 (EA1)	240
6.2.2	EA2 and Brownian path decomposition	241
6.3	Exact Algorithm 3 (EA3)	242
6.3.1	Some auxiliary results	244
6.3.2	Construction of layered Brownian bridge	248
6.3.3	Examples	250

6.4	Extensions	252
6.4.1	Multivariate diffusions	252
6.4.2	Time-inhomogeneous diffusions	253
6.5	Simulation of boundary hitting times	254
6.6	Discussion	259

List of Figures

A.1	Left: A graphical illustration of CFTP. Right: Pseudo-code	7
A.2	Left: A graphical illustration of Ro-CFTP. Right: Pseudo-code	13
A.3	Left: A graphical illustration of Dominated CFTP. Right: Pseudo-code	16
A.4	A graphical illustration of the Layered Multishift Coupler	22
1.1	A realisation of successive Ro-CFTP blocks	71
1.2	A realisation of MIC as applied to a gaussian family	77
1.3	A block of the One-Shot algorithm	81
1.4	The algorithm for coupling $TN_1(\mu, 1) _\alpha^\beta$ when $\alpha \leq \mu \leq \beta$	87
1.5	The algorithm for coupling $TN_1(\mu, 1) _\alpha^\beta$ when $\mu \leq \alpha$ or $\mu \geq \beta$	87
1.6	A realisation of our coupling method for $TN_1(\mu, 1) _{-1}^1$, $-2 \leq \mu \leq 2$	88
1.7	Left: a qq-plot for PA. Right: a qq-plot comparing PA– CSA	100
1.8	The method for drawing from $TN_1(\mu, 1) _\alpha^\beta$	106
2.1	An illustration of the sequence of increasing and decreasing sets involved in the Exact Algorithm	118
2.2	The factorisation of $\mathbb{Q}$ in terms of independent Brownian bridges	122
2.3	A case when the Exact Algorithm accepts the proposed path	126
2.4	A case when the Exact Algorithm rejects a proposed path	126
2.5	The estimated density for the distribution of X_1 from samples generated by the Euler approximation and the Exact Algorithm.	127

2.6	Exact simulation of the maximum $M_2^X = \sup\{X_t; 0 \leq t \leq 2\}$	133
2.7	Exact simulation of the hitting time $\tau_2 = \inf\{t \geq 0 : X_t = 2\}$	134
3.1	An illustration of the retrospective idea characterising EA1	147
3.2	The decomposition of a Brownian bridge at its minimum	150
3.3	An exact skeleton of the logistic growth model	158
3.4	Profile log-likelihoods from an application of the unbiased likelihood estimator on a Brownian motion	163
3.5	Profile log-likelihoods from an application of the unbiased likelihood estimator on the logistic growth model	164
4.1	The SINE and the LOG – GROWTH data sets	182
4.2	Log-likelihood estimates for the SINE and LOG – GROWTH datasets .	190
4.3	Left: The EURODOLLAR dataset. Right: Profile log-likelihood estimation with the Poisson estimator	195
4.4	Summary of MCMC results	207
5.1	An example application of SAM on the logistic growth model	226
5.2	Scheme for simulating standard Brownian bridges	231
6.1	The layers arrangement for EA3	243
6.2	A skeleton of an example SDE generated using EA3	251

List of Tables

1.1	KS tests on samples from PA and CSA	99
1.2	KS tests on samples from PA, CSA and RS	100
1.3	Draws/sec for RS, PA and CSA	101
2.1	KS tests and computing times from an application of the Exact Algorithm	128
3.1	A comparative simulation study of EA and the Euler scheme as applied to the logistic growth model	158
4.1	Summary of results from independent applications of SAM and the Poisson estimator on the logistic growth model	197
4.2	Results from an application of MCEM	202

Part I

Introduction

Review A

The CFTP Algorithm

A.1 Introduction

The Perfect Simulation idea first appeared in the seminal work of Propp and Wilson (1996). The Coupling From The Past (CFTP) algorithm introduced in that paper and its variations as presented in following papers are reviewed in the sequel. Then, many of the coupling techniques used so far in applications of CFTP are described with emphasis on methods that apply to continuous state spaces. The title Perfect Simulation is due to Kendall (1998) and is justified in Kendall and Møller (2000); it reflects the innovation of being able to simulate the equilibrium law of a Markov chain using paths of the chain without any approximation error (apart from that inherent in any computer simulation).

To determine the context, let $P : E \times \mathcal{E} \mapsto [0, 1]$ be a probability kernel on a measurable space $(E, \mathcal{E})$ and $X = \{X_t\}_0^\infty$ the corresponding discrete time Markov chain with arbitrary starting point and one-step transitions governed by P . It is assumed that X is ergodic with equilibrium law π . Thus, if P^K denotes the K -step transition kernel, $K \geq 1$, then for any $x \in E$

$$\|P^K(x, \cdot) - \pi\| = \sup_{A \in \mathcal{E}} |P^K(x, A) - \pi(A)| \rightarrow 0 \text{ as } K \rightarrow \infty, \quad (\text{A.1})$$

where $\|\cdot\|$ denotes the total variation norm. The objective is to simulate π using sample paths of X . The standard, but approximate, approach is as follows: fix X_0 and simulate $X_1, \dots, X_K$ conditionally on this X_0 for some “big enough” K . Then, after discarding some initial burn-in steps, the instances of the path are treated as correlated draws from π . The justification comes from (A.1).

In many cases qualitative results can be obtained about the type (uniform, geometric or other) of convergence to π , see e.g. Tierney (1994). In most applications however it is difficult to find explicit upper bound on the rate of convergence or when this is possible the bound is weak enough to be of no practical use (for recent advances in quantitative results on the convergence of Markov chains see e.g. Jerrum and Sinclair, 1989; Diaconis and Stroock, 1991; Meyn and Tweedie, 1994; Rosenthal, 1995; Roberts and Rosenthal, 1996). Thus one usually cannot determine a-priori a large enough K for which the distance from π will be sufficiently small. In a different approach, empirical convergence diagnostics have been developed based on exploring sample paths of X , see Cowles and Carlin (1996); Brooks and Roberts (1998) for reviews. These methods detect non-convergence only if there is significant evidence about it, so they can never reliably guarantee that convergence has been accomplished.

CFTP removes the necessity of determining the length of the burn-in. The algorithm decides by itself how much time is required before equilibrium is reached and returns an exact draw from the equilibrium law. The advantage comes at the cost of complexity though, since now trajectories of the Markov chain starting from everywhere in the state space must be simulated until they *coalesce*, i.e. merge into a single path. The precise way in which this objective translates into a practical algorithm is presented in the following sections. For the time, we will illustrate that running multiple paths and forcing coalescence requires considerable effort.

The parallel realisation of sample paths assumes a dependence structure on the transition dynamics of the Markov chain over all points of the state space.

That is, the probability measures $\{P(x, \cdot); x \in E\}$ will be *coupled*. Coupling has been a standard probabilistic technique. It can be loosely defined as the creation of dependence structure among random elements with given marginal laws, see ch.I of Lindvall (1992) for a formal definition. Its most renowned application is the proof of the ergodic theorem (A.1) by coupling a path started at x with a stationary path of X , see e.g. ch.II of Lindvall (1992). When coupling the transition laws $\{P(x, \cdot); x \in E\}$ some considerations have to be taken into account. The construction must allow for monitoring the evolution up to coalescence of sample paths initiated from everywhere in E and force coalescence itself. As a consequence, the biggest challenge to constructing efficient Perfect Simulation algorithms lies in devising coupling rules with monotone or coalescence properties. In the former case the aim is to develop upper and lower bounding paths so that coalescence can be implied when these two paths unite. In the latter case, one tries to merge paths so that it is then easier to follow their evolution. Such coupling rules are presented in the sequel.

The rest of the review is organised as follows. Section A.2 presents CFTP. Section A.3 describes its alternative format, Read-Once CFTP, which in many cases appears to be more appropriate. Section A.4 shows an extension of the CFTP protocol, Dominated CFTP, that deals with non-uniformly ergodic Markov chains. Section A.5 reviews some of the coupling techniques employed in CFTP applications. We finish with some general comments and by stating the contribution of this thesis in the Perfect Simulation context.

A.2 The CFTP protocol

CFTP requires a simultaneous simulation of paths of the Markov chain X starting from anywhere in the state space E . This translates (one only has to think of a computer code that carries out the above task) to devising a random element ξ

taking values in some space Ξ and a structure function $f : E \times \Xi \mapsto E$ such that

$$\mathbb{P}[f(x, \xi) \in dy] = P(x, dy), \quad x \in E. \quad (\text{A.2})$$

This construction lies at the heart of CFTP. Following Borovkov and Foss (1992), it generates a Stochastic Recursive Sequence representation of the Markov chain. In terms of the discussion in the introduction, it accomplishes a coupling of the one-step transition probabilities $\{P(x, \cdot); x \in E\}$. For fixed x , $f(x, \xi)$ provides a simulation method for the update of the position of X when the chain is currently at x . For fixed ξ , $f(x, \xi)$ describes the dependence structure for the one-step updates of the different sample paths. Apparently, there can be various choices for f valid under the requirement (A.2).

Let $\{\xi_t\}_{t=-\infty}^{\infty}$ be independent copies of ξ ; we simplify $f(\cdot, \xi_t)$ to $f_t(\cdot)$. Using f one obtains multi-step updates of paths between the time instances s and t , by setting

$$F_s^t(\cdot) := f_{t-1} \circ \cdots \circ f_{s+1} \circ f_s(\cdot), \quad s \leq t, \quad (\text{A.3})$$

under the convention that F_t^t , for any integer t , is the identity map. Note that s, t will be allowed to be negative, i.e. we will consider sample paths started at negative times. The random function f must be constructed in a way that

$$\text{there exists a fixed } T > 0 \text{ such that } \mathbb{P}[F_0^T(E) \text{ is a singleton}] > 0, \quad (\text{A.4})$$

where, for any $t \geq 0$, $F_0^t(E)$ is the random image $\{F_0^t(x); x \in E\}$.

For simplicity, CFTP will be described for the case when E is finite; thus, after relabelling if necessary, $E = \{1, 2, \dots, n\}$. The general case follows in the same lines. Assume that $P = \{P_{ij}\}$ is the one-step transition matrix. For E as above, (A.2) specialises to

$$\mathbb{P}[f(i, \xi) = j] = P_{ij}, \quad i, j \in E.$$

From the ergodicity of X and the finiteness of the state space, condition (A.4) will be always satisfied in the simple case when f is obtained using independent updates for each $i \in E$, see Theorem 2 of Propp and Wilson (1996).

The standard method for generating approximate draws from π is to consider the *forward* path $F_0^t(i)$ for a given starting point i and a sufficiently big t . All Markov chain Monte Carlo (MCMC) algorithms are examples of this approach. In a similar spirit, but just one step away from CFTP, is the work of Johnson (1996). He considers not a single but multiple coupled paths starting from various positions of the state space to obtain empirical results about the rate of convergence to the stationary law after examining the time needed for the paths to merge. Johnson (1996) also notes that the forward coupling time

$$\tau_f := \inf\{t > 0 : F_0^t(i) = F_0^t(j) \text{ for all } i, j \in E\},$$

which under (A.4) is a.s. finite, is not a correct stopping rule for simulating π , i.e. in general the singleton $F_0^{\tau_f}(E) = \{F_0^{\tau_f}(i); i \in E\}$ is not a draw from π . A time-reversal argument will now lead to CFTP.

Consider the *backward* coupling time

$$\tau_b := \inf\{t > 0 : F_{-t}^0(i) = F_{-t}^0(j) \text{ for all } i, j \in E\}, \quad (\text{A.5})$$

which is equal in distribution to τ_f , see Foss and Tweedie (1998), thus a.s. finite under (A.4). The backward and forward maps, F_{-t}^0 and F_0^t respectively, are fundamentally different. As t increases F_{-t}^0 evolves paths from deeper into the past up to time 0 while F_0^t adds additional steps to a typical forward in time path simulation. The most direct consequence of moving the starting point of the paths further into the past instead of moving their ending point into the future is the CFTP protocol itself.

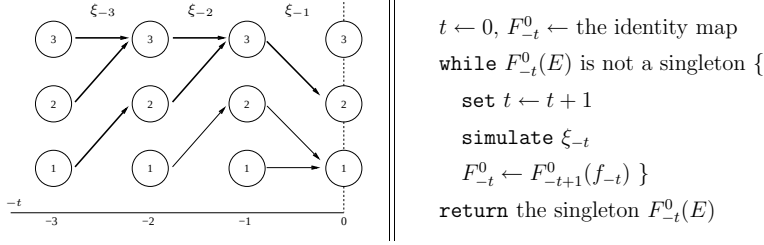


Figure A.1: Left panel: A graphical illustration of CFTP. In this case $E = \{1, 2, 3\}$. The arrows show a realisation of the one-step updates $f(\cdot, \xi_t)$ for $t = -1, -2, -3$; e.g. $f(3, \xi_{-1}) = 2$ and $f(1, \xi_{-2}) = 2$. The thick arrows show that coalescence occurs after travelling $\tau_b = 3$ steps into the past, with output $F_{-\tau_b}^0(E) = \{2\}$. Right panel: Pseudo-code for CFTP.

Theorem A.1 (CFTP) *The singleton $F_{-\tau_b}^0(E) = \{F_{-\tau_b}^0(i); i \in E\}$ is distributed according to π .*

This is Theorem 2 of Propp and Wilson (1996). The intuition is clear. Consider a path started sufficiently deep into the past (say from $-\infty$) to have reached stationarity up to time $-\tau_b$. Then necessarily the path will be at $F_{-\tau_b}^0(E)$ at the fixed time 0, which is thus a draw from the stationary law.

Fig.A.1 shows a graphical illustration of CFTP and the pseudo-code for the implementation of CFTP on a computing system. The code exploits the identity $F_{-t}^0(\cdot) = F_{-t+1}^0(f_{-t}(\cdot))$ determined from (A.3) and by storing the map $i \mapsto F_{-t+1}^0(i)$, it constructs F_{-t}^0 after simulating only the single additional step f_{-t} . It thus avoid running paths all the way up to time 0 whenever a time further into the past is accessed.

Even under the above consideration the computational effort for simulating and storing sample paths moving anywhere in the state space can be substantial (and sometimes prohibiting) when the state space is large; the icing model, for instance, has state space of size 2^N , where N is the number of vertices in the model. Similar efficiency problems can appear when the coupling time τ_b is large.

Note here that, in general, the output of CFTP is not independent of the run-

ning time of the algorithm. Thus, if a long run is aborted before it returns a draw then the final sample will be biased in favour of the states preferred by small τ_b 's. But, since τ_b is unbounded, there will always be a positive probability of aborting a CFTP run in any case when a specific time horizon for collecting a sample from π has been forced. This subtle point is termed “user-impatience” bias in Propp and Wilson (1998a) where it is also discussed in detail. Fill (1998) introduced an “interruptible” algorithm that avoids the user-impatience bias. The algorithm is based on rejection sampling and follows the same principle of coupling paths of a Markov chain to draw from its equilibrium law. Its applicability, however, is much more restricted compared to CFTP.

A.2.1 CFTP and stochastic monotonicity

CFTP is mostly efficient when applied to Markov chains with monotonicity properties. Assume that E admits a partial ordering ($\preceq$) with minimal and maximal elements $\hat{0}, \hat{1}$ respectively and that $f(\cdot, \xi)$ preserves the space ordering, that is, w.p.1

$$f(i, \xi) \preceq f(j, \xi) \text{ if } i \preceq j.$$

Then, for any $0 \leq s \leq t$ and $i \in E$, w.p.1

$$F_{-t}^{-s}(\hat{0}) \preceq F_{-t}^{-s}(i) \preceq F_{-t}^{-s}(\hat{1}),$$

so all paths will be constrained between an upper and a lower path. Consequently, in this case

$$\tau_b = \inf\{t > 0 : F_{-t}^0(\hat{0}) = F_{-t}^0(\hat{1})\}, \quad (\text{A.6})$$

and since $F_{-\tau_b}^0(E) = F_{-\tau_b}^0(\hat{0}) = F_{-\tau_b}^0(\hat{1})$, Theorem A.1 can be applied by simulating only pairs of paths started at time $-t$ from $\hat{0}, \hat{1}$ and constructed up to time 0, for $t = 1, \dots, \tau_b$. Note that the random variables $F_{-t}^0(i)$, for all $t < 0$, $i \in E$,

are defined in (A.3) over a single sequence of $\{\xi_t\}_{-\infty}^{-1}$. Thus, the same realisations $\xi_{-1}, \xi_{-2}, \dots$, must be used every time that paths coming from further into the past arrive at the time instances $-1, -2, \dots$, respectively. It should also be mentioned that going only one step back at a time can be unnecessarily costly. Propp and Wilson (1996) argue that an efficient policy is to start paths at the times $-2^k t_0$, for $k = 0, 1, \dots$, with some chosen initial step t_0 . This scheme has been followed by almost all CFTP algorithms (monotone or not) that appeared later in the literature.

Propp and Wilson (1996) show that the backward coupling time (A.6) is nearly the best possible under any coupling of the monotone Markov chain since

$$\mathbb{E}[\tau_b] \leq 2\tau_{\text{mix}}(1 + \log l)$$

where τ_{mix} is the *mixing time* of the Markov chain, defined as the smallest $K \geq 1$ such that $\sup_{i,j \in E} \|\mathbb{P}[X_K \in \cdot | X_0 = i] - \mathbb{P}[X_K \in \cdot | X_0 = j]\| \leq 1/e$, and l is the largest number of elements that can exist in a totally ordered subset of E . Thus, to quote Propp and Wilson (1996), “if the Markov chain is rapidly mixing then it is rapidly coupling”.

Certainly, monotone coupling is possible only under a specific structure for the transition matrix. A necessary condition is that the one-step transition dynamics are stochastically ordered, i.e. for any $i \preceq j$ and any increasing $g : E \mapsto \mathbf{R}$, $\mathbb{E}[g(X_1) | X_0 = i] \leq \mathbb{E}[g(X_1) | X_0 = j]$, see ch.IV of Lindvall (1992); g is increasing when $i \preceq j$ implies $g(i) \leq g(j)$. Sufficient conditions for the existence of a monotone coupling in the case of a finite state space are given in Fill and Machida (1999). These results are mostly of theoretical nature; in cases of applications of monotone CFTP in the literature (see the references in the following section) the Markov chains have provided natural monotone update rules.

A.2.2 Applications of CFTP

Applications of CFTP to finite state spaces include the random cluster and tiling models using monotone CFTP (Propp and Wilson, 1996), random spanning trees (Propp and Wilson, 1998b), and the hard-core and random cluster model (for different parameter values than above) using *anti*-monotone CFTP (Häggström and Nelander, 1998). The general CFTP idea, as already described, extends without any modification to arbitrary state spaces. In continuous state spaces though, generating update rules that preserve the stationary law, force coalescence and allow monitoring the evolution of uncountably many paths up to coalescence will necessarily involve sophisticated coupling techniques. For example applications of CFTP in stochastic geometry and point processes see Kendall (1998); Kendall and Thönnies (1999); Häggström et al. (1999); Kendall and Møller (2000). For CFTP applied to MCMC samplers on Euclidean subspaces see e.g. Murdoch and Green (1998b); Green and Murdoch (1999); Møller (1999); Wilson (2000c); Mira et al. (2001); Corcoran and Tweedie (2002).

Some of the coupling techniques of these papers are presented in Section A.5. In all up-to-date applications of CFTP to continuous state spaces, the substantial algebraic effort to build an algorithm monitored in each case on a specific probability model (see Section A.5 for more details) illustrates that CFTP algorithms are still far from being generally applicable or automatic.

A.3 Reversing the time: Read-Once CFTP

Wilson (2000b) introduced a version of CFTP that evolves forwards in time. The algorithm is termed Read-Once CFTP (Ro-CFTP) since, unlike the original CFTP, it does not require the storage of currently realised random variates or maps to be used by subsequent paths of the Markov chain, instead it generates paths following a read-once stream of random elements. Ro-CFTP appears to be more appropriate

than CFTP in cases of complex algorithms on continuous state spaces; its advantages are briefly examined at the end of the section. Before that, the algorithm is presented under the discrete state space setting of Section A.2.

It is now necessary that the Markov chain and the rule $f(\cdot, \xi)$ for the simulation of coupled paths of it considered in Section A.2 are such that (A.4) holds for $T = 1$, that is

$$P[f(i, \xi) = f(j, \xi) \text{ for all } i, j \in E] > 0. \quad (\text{A.7})$$

The update rule with the one-step coalescence property (A.7) could be thought of as the composition of preliminary update rules for which (A.4) holds for some $T > 1$. Assume now a CFTP setting that considers the backward coupling time

$$\tau_r := \inf\{t > 0 : f(i, \xi_{-t}) = f(j, \xi_{-t}) \text{ for all } i, j \in E\},$$

and returns the singleton $F_{-\tau_r}^0(E)$. Note that $\tau_r \geq \tau_b$, for τ_b defined in (A.5), and $F_{-\tau_r}^0(E) = F_{-\tau_b}^0(E)$, thus the singleton $F_{-\tau_r}^0(E)$ is a draw from π .

Let ρ denote the coalescence probability in (A.7). Clearly, τ_r follows a Geometric distribution with parameter ρ . Given τ_r , the seeds $\xi_{-\tau_r+1}, \xi_{-\tau_r+2}, \dots, \xi_{-1}$ are iid and conditioned not to induce coalescence while $\xi_{-\tau_r}$ is independent of $\xi_{-\tau_r+1}, \xi_{-\tau_r+2}, \dots, \xi_{-1}$ and conditioned to yield coalescence. Thus, if ρ was known explicitly and draws from the distribution of ξ conditionally on coalescence or not coalescence for the update rule $f(\cdot, \xi)$ were available then one could readily simulate $F_{-\tau_r}^0(E)$ running a single path from $-\tau_r$ to 0. But, there is no reason now for this path to evolve in negative times. One can equivalently simulate ξ_0 conditionally on $f(\cdot, \xi_0)$ being coalescent, then, independently, $\xi_1, \dots, \xi_{\tau_r-1}$ conditionally on the relevant updates being non-coalescent and output the singleton $F_0^{\tau_r}(E)$ as an exact draw from π .

The difficulty in this heuristic approach is that the above path cannot be directly simulated. It can, however, be simulated indirectly. Consider the maps

$f(\cdot, \xi_t)$, $t \geq 0$, and the times of the first and second coalescence $\tau_1 = \inf\{t \geq 0 : f(i, \xi_t) = f(j, \xi_t) \text{ for all } i, j\}$ and $\tau_2 = \inf\{t > \tau_1 : f(i, \xi_t) = f(j, \xi_t) \text{ for all } i, j\}$ respectively. The seeds $\xi_{\tau_1}, \xi_{\tau_1+1}, \dots, \xi_{\tau_2-1}$ are of number following a geometric distribution with parameter ρ ; ξ_{τ_1} is a copy of ξ conditioned to force coalescence, and $\xi_{\tau_1+1}, \dots, \xi_{\tau_2-1}$ are independent copies of ξ conditioned not to bring coalescence. In the light of the above discussion, $F_{\tau_1}^{\tau_2}(E)$ will be a draw from π . The seed ξ_{τ_2} that forces coalescence, served only as a stopping rule without any other information about it used at obtaining $F_{\tau_1}^{\tau_2}(E)$ thus it can be utilised as the initial point for another sequence $\xi_{\tau_2}, \xi_{\tau_2+1}, \dots, \xi_{\tau_3-1}$, where τ_3 is defined in the obvious way, and so on.

Ro-CFTP has now been heuristically demonstrated; the above discussion is in line with the description of Wilson (2000b). Formally, Ro-CFTP develops as follows.

Theorem A.2 (Read-Once CFTP) *The sequence of random times*

$$\begin{aligned} \tau_1 &= \inf\{t \geq 0 : f(i, \xi_t) = f(j, \xi_t) \text{ for all } i, j\}, \\ \tau_i &= \inf\{t > \tau_{i-1} : f(i, \xi_t) = f(j, \xi_t) \text{ for all } i, j\}, \quad i > 1, \end{aligned}$$

is such that the singletons $F_{\tau_i}^{\tau_{i+1}}(E)$, $i \geq 1$, are iid draws from π .

For a rigorous proof see Breyer and Roberts (2001). The left panel in Fig.A.2 provides a graphical illustration of Ro-CFTP. The figure shows the realisation of the random maps $f(\cdot, \xi_t)$, for $t = 0, \dots, \tau_2$ and, circled, the singleton $F_{\tau_1}^{\tau_2}(E)$ which corresponds to an exact draw from π .

A computer code implementation of Ro-CFTP would require a binary flag for each realisation of $f(\cdot, \xi_t)$, $t \geq 0$, to distinguish between coalescence and non-coalescence and the monitoring of the evolution of the path initiated by the singleton output of the first coalescent map $f(\cdot, \xi_{\tau_1})$. The instances of this path one step before each subsequent coalescence correspond precisely to the singletons $F_{\tau_i}^{\tau_{i+1}}(E)$

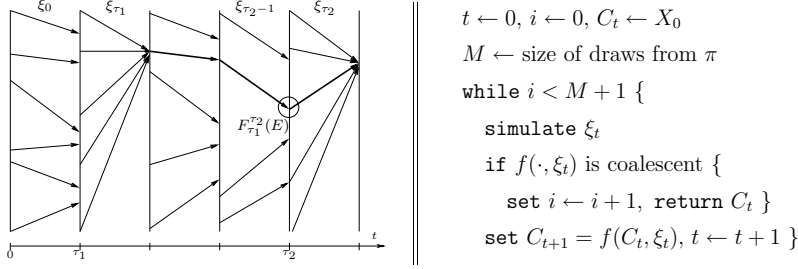


Figure A.2: Left panel: A graphical illustration of Ro-CFTP. The coalescent maps $f(\cdot, \xi_{\tau_1})$ and $f(\cdot, \xi_{\tau_2})$ determine the singleton $F_{\tau_1}^{\tau_2}(E)$ which is a draw from π . Right panel: Pseudo-code for getting M draws from π using Ro-CFTP. The algorithm requires $M + 1$ coalescent f -maps and returns $M + 1$ values from which the last M constitute the required sample from π .

and thus are independent draws from π . The right panel in Fig.A.2 presents the pseudo-code for this algorithm. We have denoted by C_t the path determined above, which for simplifying the code we have allowed to commence at time 0 (and not at the time of the first coalescence as discussed above) from some arbitrary $X_0 \in E$.

For applications of Ro-CFTP see Wilson (2000b) who simulates some locally stable point processes and Breyer and Roberts (2001) for an algorithm on a hierarchical poisson/gamma model. Compared with CFTP, Ro-CFTP has the obvious advantage that running paths in the natural time direction simplifies coding. Also, seeds or maps are only assessed once and are then discarded thus there is no need of storing seeds or synthesising maps coming from deeper into the past with maps already realised which could require complex calculations and separate code.

A.4 Dominated CFTP

Foss and Tweedie (1998) showed that the existence of an a.s. finite backward coupling time for CFTP is equivalent to the Markov chain under consideration being uniformly ergodic. Though uniform ergodicity will always be the case for an ergodic chain on a finite state space, one usually encounters weaker forms of

convergence on unbounded spaces. Kendall (1998) presented a modification of the original CFTP protocol, termed Dominated CFTP (and sometimes referred to as CIAFTP - coupling *into* and from the past) that extends considerably the applicability of CFTP. Kendall (2004) demonstrates that there exists a version of Dominated CFTP (though not necessarily identifiable in practice) for any case of a geometrically ergodic Markov chain. In general, Dominated CFTP can be thought of as CFTP with upper and lower states artificially obtained in terms of a pair of dominating processes. For a detailed presentation of the algorithm see Kendall and Møller (2000).

The state space is now assumed to be partially ordered, w.r.t. a relation $(\preceq)$, but otherwise arbitrary. In accordance to Kendall and Møller (2000), the algorithm will be described for the simpler case when there exists a minimal element $\hat{0}$. For the general case (which is considerably harder) with an application see Montana (2002). We now use the general definitions of the update rules (A.2), (A.3) that consider arbitrary state spaces.

Assume that $D = \{D_t\}$ is an ergodic Markov chain on E with stationary law π_d that stochastically dominates X in a way that one can obtain an update rule f_d for D coupled with f so that w.p.1

$$f(x, \xi) \preceq f_d(y, \xi) \text{ if } x \preceq y. \quad (\text{A.8})$$

The algorithm begins by observing at time 0 the location of a path of D started at $-\infty$, thus it draws $D_0 \sim \pi_d$. A subtle simulation method is now required that for a given D_0 travels *into* the past up to any requested time instance $-t < 0$ to simulate $D_{-1}, \dots, D_{-t}$ using the backward dynamics of D . Analytically, if $P_d(x, dy)$ is the transition kernel of D , then $D_{-s}|D_{-s+1} = x$ must be simulated according to the backward kernel

$$\tilde{P}_d(x, dy) := \frac{P_d(y, dx)}{\pi(dx)} \cdot \pi(dy). \quad (\text{A.9})$$

The method must then travel *from* the past back to time 0 to deduce the seeds $\xi_{-t}, \dots, \xi_{-1}$ that would have generated the path $D_{-t+1}, \dots, D_{-1}, D_0$ under the update rule f_d had they been given D_{-t} as a starting point.

Assume that f is monotone, that is w.p.1

$$f(x, \xi) \preceq f(y, \xi) \text{ if } x \preceq y. \quad (\text{A.10})$$

Under this condition, and (A.8), one obtains that for any $0 \leq s \leq t$, w.p.1

$$F_{-t}^{-s}(\hat{0}) \preceq F_{-t}^{-s}(x) \preceq F_{-t}^{-s}(D_{-t}) \preceq D_{-s} \text{ if } \hat{0} \preceq x \preceq D_{-t}, \quad (\text{A.11})$$

where for a fixed t , the variables $\xi_{-t}, \dots, \xi_{-1}$ used for these paths are obtained conditionally on $D_0, D_{-1}, \dots, D_{-t}$ in the CIAFTP way sketched above. Furthermore, (A.8) and (A.10) imply that for $0 \leq s \leq t < u$, w.p.1

$$F_{-t}^{-s}(\hat{0}) \preceq F_{-u}^{-s}(\hat{0}) \preceq F_{-u}^{-s}(D_{-u}) \preceq F_{-t}^{-s}(D_{-t}). \quad (\text{A.12})$$

Equations (A.11) and (A.12) are special cases of what Kendall and Møller (2000) call the “sandwiching” and “funnelling” property respectively. Kendall and Møller (2000) describe a more general algorithm that does not assume the monotone rule (A.10); in their algorithm, the bounding paths $s \mapsto F_{-t}^{-s}(\hat{0})$, $s \mapsto F_{-t}^{-s}(D_{-t})$, used in (A.11) and (A.12) for various initial instances $-t$, need not be paths of X but are defined in an abstract way that guarantees precisely a sandwiching and funnelling property. Consider the backward coupling time

$$\tau_d = \inf\{t > 0 : F_{-t}^0(\hat{0}) = F_{-t}^0(D_{-t})\}.$$

The similarity to (A.6) is striking. One can actually think of the monotone CFTP algorithm of Section A.2.1 as a special case of Dominated CFTP as described

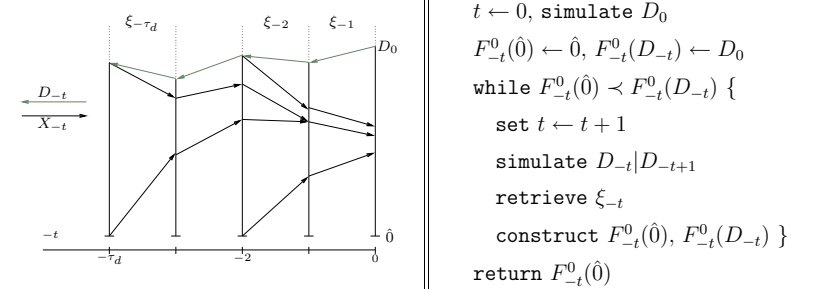


Figure A.3: Left panel: A graphical illustration of Dominated CFTP. The path of the dominating chain D travels into the past and determines the upper bound $t \mapsto D_{-t}$ on the domain of the target chain X . The graph shows the realisation of the paths $s \mapsto F_{-t}^{-s}(D_{-t})$ and $s \mapsto F_{-t}^{-s}(\hat{0})$ for $t = 2$ and $t = \tau_d (= 4)$. Right panel: Pseudo-code for Dominated CFTP.

here with the process D being constant and equal to the upper bound $\hat{1}$ assumed therein.

It is necessary that τ_d is a.s. finite; Kendall and Møller (2000) obtain a.s. finiteness under the assumption that $\hat{0}$ is an ergodic atom for D , i.e. that for any initial point D_0 , D will visit 0 w.p.1 in finite positive time.

Theorem A.3 (Dominated CFTP) $F_{-\tau_d}^0(\hat{0})$ is a draw from π .

For the proof see Kendall and Møller (2000). For the intuition behind Dominated CFTP, imagine a path of D and a path of X started at $-\infty$, the former path started higher than the latter, and evolved according to the update rules f_d, f respectively. From (A.8), $X_{-t} \preceq D_{-t}$ for all $t \geq 0$. The objective is to locate $X_0 \sim \pi$. At time $-\tau_d$ all the possible states for $X_{-\tau_d}$, i.e. those that are found lower than $D_{-\tau_d}$, lead to the unique state $F_{-\tau_d}^0(\hat{0})$ at time 0 which therefore coincides with X_0 .

The left panel in Fig.A.3 gives a graphical illustration of Dominated CFTP; the algorithm determines an artificial upper bound for the state space. The simple monotone CFTP protocol of Section A.2.1 is then applied on the restricted domain. The right panel in Fig.A.3 presents the pseudo-code for the Dominated CFTP algorithm as described above.

Dominated CFTP exploits a *horizontal* backward coupling time; for a given state (here $\hat{0}$) a random time in the past is found for which paths started at $\hat{0}$ at any time instance before this random time, finish at time 0 at the same state. In contrast, CFTP uses a *vertical* backward coupling time; a random time in the past is found so that simulations from that time up to time 0 project all the state space onto a singleton. Since any vertical coupling time is always an horizontal one, Dominated CFTP appears also in this sense as an extension of CFTP.

Kendall (1998) presented the first application of Dominated CFTP for the case of an area-interaction point process. Then, Kendall and Møller (2000) extended the algorithm to general locally stable point processes. See also Mira et al. (2001) and Green and Murdoch (1999) for applications involving the MCMC slice sampler and random walk Metropolis respectively. In brief, Dominated CFTP is a sophisticated technique that requires complex coupling constructions. Finding a Markov chain with known stationary law that stochastically dominates the target chain is a demanding requirement. Running the dominating chain into the past under the backward kernel (A.9) and then retrieving the forward random seeds will typically involve intricate calculations.

A.5 Coupling methods

The importance of coupling $\{P(x, \cdot); x \in E\}$, that is devising a structure function f with the property (A.2), in a way that facilitates CFTP purposes (i.e. f being monotone or forcing coalescence) has by now been emphasised. In the sequel, some of the methods for coupling an one-step transition kernel developed in the literature are reviewed; references about the complete CFTP algorithms that incorporate these coupling constructions will be given.

Many of the methods are using transition densities, thus in what follows $p(x, \cdot)$, $x \in E$, denotes the density (assuming it exists) of $P(x, \cdot)$ w.r.t. some reference

measure λ .

A.5.1 Multigamma coupling

Multigamma coupling derives from the γ -coupling of pairs of probability measures; see e.g. ch.I of Lindvall (1992). It also relates with the splitting method of Athreya and Ney (1978) and Nummelin (1978). In the context of CFTP it has been introduced by Murdoch and Green (1998b). Multigamma coupling develops as follows. Assume that there exists a probability measure ν on $(E, \mathcal{E})$ and a real $0 < \delta < 1$ such that

$$P(x, A) \geq \delta \cdot \nu(A), \quad x \in E, A \in \mathcal{E}.$$

This one-step minorisation condition implies that P is uniformly ergodic, see Theorem 16.0.2 of Meyn and Tweedie (1993). Consider also the residual probability kernel defined as $Q(x, \cdot) = [P(x, \cdot) - \delta \nu(\cdot)] / (1 - \delta)$. Then,

$$P(x, \cdot) = \delta \nu(\cdot) + (1 - \delta) Q(x, \cdot). \quad (\text{A.13})$$

Let f_q be an update rule for Q , that is $f_q = f_q(x, \xi_q)$ for some random element ξ_q , with $P[f_q(x, \xi_q) \in dy] = Q(x, dy)$, $x \in E$. Let $\xi_G = (\xi_q, U, V)$, where $U \sim \text{Un}(0, 1)$ and $V \sim \nu$. Consider the random function

$$f_G(x, \xi_G) = \begin{cases} V, & \text{if } U < \delta \\ f_q(x, \xi_q), & \text{otherwise} \end{cases}, \quad x \in E.$$

From (A.13), clearly f_G couples $P(x, \cdot)$. Notice that $P[f_G \text{ is coalescent}] \geq \delta$.

Naive implementations of multigamma coupling on large state spaces can be extremely inefficient; Murdoch and Green (1998b) note that had they applied multigamma coupling as described so far to the example poisson/gamma hierarchical model they consider, they would have to use $\delta \approx 10^{-24}$. Instead, they develop a simple improvement of it, termed partitioned multigamma coupling, by

partitioning the state space and establishing separate minorisation conditions on each piece of the partition; see Murdoch and Green (1998b) for more details. Note also that minorisation conditions require explicit knowledge of the probability kernel (not just up to a constant of proportionality) and this restricts the applicability of the method.

A.5.2 Rejection coupling

Murdoch and Green (1998b) suggested also a rejection sampling based coupling method that does not require normalising constants. Assume that the transition density of the given Markov chain can be written as $p(x, \cdot) = c(x) p_u(x, \cdot)$ for some positive constant $c = c(x)$ and an unnormalised density $p_u(x, \cdot)$ and that

$$p_u(x, y) \leq r(y), \quad x, y \in E, \quad (\text{A.14})$$

for a λ -integrable $r : E \mapsto \mathbf{R}$. Let $\mathfrak{U} = (U_1, U_2, \dots)$, $\mathfrak{V} = (V_1, V_2, \dots)$ be two sequences of iid random elements with $U_1 \sim \text{Un}(0, 1)$ and $V_1 \sim r(\cdot) / \int_E r(y) \lambda(dy)$. Consider the compound element $\xi_R = (\mathfrak{U}, \mathfrak{V})$ and the random function

$$f_R(x, \xi_R) = V_{\tau_x}, \quad x \in E, \quad (\text{A.15})$$

where

$$\tau_x = \inf \{i \geq 1 : U_i \leq p_u(x, V_i) / r(V_i)\}.$$

Simple rejection sampling theory can certify that f_R couples $\{P(x, \cdot); x \in E\}$. Murdoch and Green (1998b) obtain a sufficient condition for f_R to be determined after an a.s. finite number of proposals by requiring that there exists a non-negative $l : E \mapsto \mathbf{R}$ with $\int_E l(y) \lambda(dy) > 0$ such that

$$l(y) \leq p_u(x, y), \quad x, y \in E.$$

In practice, (A.15) is implemented by finding a subset of the state space attracted by each successive proposal until the union of these sets cover all the state space. See Murdoch and Green (1998b) for an application to the poisson/gamma hierarchical model also referenced in Section A.5.1.

The appealing characteristic of rejection coupling is that it projects the state space to a discrete collection of points. Its applicability is restricted by (A.14) which is difficult to obtain in the case of an unbounded state space.

A.5.3 Catalytic coupling

The coupling techniques presented so far build on the specific properties of the probability kernel under consideration. Breyer and Roberts (2001) presented a method which, at least at its definition, is automatic and applies under minimal assumptions (namely, the existence of a density for $P(x, \cdot)$). It is based on the Independence sampler MCMC method.

Let $f_0(x, \xi_0)$, for some random element ξ_0 , be a preliminary coupling of the transition laws $\{P(x, \cdot); x \in E\}$. Assume that q is a density (possibly unnormalised) on E w.r.t. the reference measure λ and that $Y \sim q$. Let $\xi_C = (\xi_0, Y, U)$ where $U \sim \text{Un}(0, 1)$ and consider the random function

$$f_C(x, \xi_C) = \begin{cases} Y, & \text{if } p(x, f_0(x, \xi_0)) q(Y) U \leq p(x, Y) q(f_0(x, \xi_0)) \\ f_0(x, \xi_0), & \text{otherwise} \end{cases}, \quad x \in E. \quad (\text{A.16})$$

That f_C couples $\{P(x, \cdot); x \in E\}$ is directly implied by the Independence sampler idea; f_C chooses between a draw from $p(x, \cdot)$ and an independent proposal with an accept-reject mechanism that preserves $p(x, \cdot)$. For a formal demonstration of the correctness of (A.16) see Breyer and Roberts (2001).

The initial coupling f_0 required by (A.16) is many times easy to obtain, as for example in the case of a Gibbs sampler. The modified coupling f_C is more relevant

to CFTP purposes since it projects parts of the state space to single points. For a proposal density, Breyer and Roberts (2001) suggest using $q = p(x^*, \cdot)$ for some $x^* \in E$, or more generally several densities determined in this way for a collection of points scattered around E ; each of these densities uses the effect of the previous ones on a given initial coupling of $\{P(x, \cdot); x \in E\}$ as the input coupling which in its turn will modify. Intuitively, by choosing a proposal density $q = p(x^*, \cdot)$ one manages to “stimulate” the space around x^* in the sense that, for sufficiently regular E and $p(x, \cdot)$, a continuum of points around x^* will accept the move to $Y \sim q = p(x^*, \cdot)$ (to see that, note that for x^* itself, $f_c(x^*, \xi_c) = Y$ w.p.1). This argument should also justify the title “Catalytic Coupling”.

Breyer and Roberts (2001) apply catalytic coupling to the same poisson/gamma model used by Murdoch and Green (1998b). They employ several proposal densities in the way described above in order to obtain a high probability for the state space to be projected onto a finite number of points. Note that some cumbersome algebraic calculations will be usually required when determining the set of points that accept (or not) the independent proposals. Also the generality of the method comes with the cost of not taking advantage of convenient characteristics (e.g. monotonicity) of the specific structure of the given probability kernel.

A.5.4 Layered multishift coupling

Wilson (2000c) introduced a monotone coupling for univariate uniform random variables restricted on shifted intervals of the same length. Using slice sampling, the algorithm can then be extended to cover more general families of distributions.

Consider the family of uniform distributions $\mathcal{U}_{L,R} = \{\text{Un}(L+s, R+s); s \in \mathbf{R}\}$ where $L < R$. A simple way to couple the members of this family is by rearranging their domains, see Fig.A.4. Furnishing the simple algorithm of the figure with a mathematical formula results in the following coupling of $\mathcal{U}_{L,R}$ (using

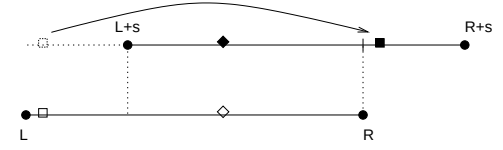


Figure A.4: In this case $0 < s < R - L$. If $U \sim \text{Un}(L, R)$ is in $[L + s, R]$, whence relabelled as $\diamond$, return $\diamond = \diamond$; else if $U \in [L, L + s)$, whence relabelled as $\square$, return $\blacksquare = \square + (R - L)$. The output will be a draw from $\text{Un}(L + s, R + s)$.

a $U \sim \text{Un}(L, R)$),

$$f_u(s, U) = \left\lfloor \frac{s + R - U}{R - L} \right\rfloor (R - L) + U, \quad s \in \mathbf{R}. \quad (\text{A.17})$$

Clearly, f_u is increasing in s . Also, its image is discrete; for a given integer $k \geq 1$, the set $\{f_u(s, U); s \in [0, k(R - L)]\}$ consists of $k + 1$ points.

Coupling of uniform laws can be utilised for the coupling of more general distributions. Consider for example the laws $\{\mathcal{N}(x, 1); x \in \mathbf{R}\}$. Fix some x and let $(U_x, V) \sim \text{Un}[(u, v) \in \mathbf{R}^2 : v \leq \mathcal{N}_{x,1}(u)]$. Then

$$U_x | V = v \sim \text{Un}[A_{x,v}], \text{ for } A_{x,v} = \{u \in \mathbf{R} : v \leq \mathcal{N}_{x,1}(u)\}.$$

Note that marginally $U_x \sim \mathcal{N}(x, 1)$ and for a given v the length of the interval $A_{x,v}$ does not depend on x . Thus, one can couple $\{\mathcal{N}(x, 1); x \in \mathbf{R}\}$ by using a common draw v of the height V from its marginal distribution (easily obtained by drawing $U_x \sim \mathcal{N}(x, 1)$ and then $V | U_x = u \sim \text{Un}[0, \mathcal{N}_{x,1}(u)]$ for some x) and then applying (A.17) for L, R such that $[L, R] = A_{0,v}$ and $s = x$.

For x restricted in a bounded interval this method yields a coupling for which the number of elements of its image will be a.s. finite but not bounded. To force a bound, Wilson (2000c) exploits a reflection property of the gaussian density which allows him to call on (A.17) for a random length $R - L$ lower bounded by a positive real.

Wilson (2000c) also notes that the slice sampling motivated layered multishift coupling applies in the same way to any univariate, unimodal family of shifted distributions as long as one can draw from a density of this family (and thus draw the height V above) for which also knows its mode and the inverse functions for its two branches opposite to the mode (and thus find the intervals $A_{x,v}$ above). He implements his method to a Gibbs sampler on the auto-normal distribution. Clearly, the applicability of the layered multishift coupling is restricted to one-dimensional families of distributions generated by shifting a given base distribution; even in this case the method requires substantial information on the structure of the density of these distributions.

A.6 Some general comments

The simplicity of the CFTP idea has fascinated researchers working on the simulation of stochastic systems. After the introduction of the algorithm in Propp and Wilson (1996), a large number of papers presenting applications of CFTP to stochastic geometry, point processes, hierarchical model simulation and other probability models and extensions of the initial CFTP idea have appeared. An up-to-date list of articles related to CFTP is maintained by David Wilson and is publicly available at the URL

www.dimacs.rutgers.edu/~dbwilson/exact/

The overwhelming embracing of the new idea by several applied probability researchers and the subsequent energetic activity on extending its applicability has by now provided a clearer picture for the potential and the limitations of CFTP. The initial perception that the algorithm would be rapidly available for routine MCMC applications on continuous or unbounded state spaces is now replaced by the acknowledgement that considerable effort is still required before generically applicable CFTP algorithms are made available.

CFTP is mostly efficient in the case of finite state space models, where it appears to be competitive (and many times superior) to standard approximate MCMC applications. The development of the CFTP algorithm becomes harder in the case of the continuous and typically high-dimensional state spaces arising in Bayesian problems. Almost all coupling methods used so far in such cases (like for instance the methods presented in Section A.5) are model-specific and require intricate algebraic calculations and the development of highly complicated computer code. It is clear that for Perfect Simulation to be established as a standard method for the MCMC practitioner, coupling techniques of general applicability must be devised. Also, complete algorithms implementable on a wide range of models requiring a minimum tuning effort from the user's side must be developed.

A.7 Our contribution

In Article 1, we present a general framework (termed One-Shot CFTP) for CFTP applications. Our algorithm runs coupled paths until they get sufficiently close. It then applies a separate coupling method to force coalescence. We apply our method to a family of multidimensional truncated gaussian distributions and show that, with minimal modifications, the same algorithm can apply to more general families of distributions.

Our algorithm combines coupling techniques for continuous state spaces that have already appeared in the literature, but also makes use of a new coupling method that we introduce in Article 1. The method, termed the Monotone Independence Coupler, applies to one-dimensional stochastically ordered probability laws. It can then be implemented under the weak condition that the stochastic ordering can be realised and the assumption of the existence of a probability density for the laws under consideration. Once these conditions are met, the development of the computer code for our coupler requires simply simulation from an arbitrary

member of the probability laws.

Review B

Approximate Simulation Methods for SDEs

B.1 Introduction

We review the most common approximate time discretisation schemes for the simulation of diffusions. For each method we describe the corresponding systematic error. The main reference for the material presented in the sequel is Kloeden and Platen (1992). The objective diffusion is assumed to solve the stochastic differential equation (SDE)

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad t \in [0, T], \quad X_0 = x_0, \quad (\text{B.1})$$

for a scalar Brownian motion $\{W_t; t \in [0, T]\}$ and drift and coefficient functions $b : \mathbf{R} \mapsto \mathbf{R}$ and $\sigma : \mathbf{R} \mapsto \mathbf{R}$ respectively.

Methods for the numerical solution of diffusions are necessary since, except for a few trivial cases, SDEs used in applications cannot be explicitly solved. Simulations from the diffusion can subsequently provide Monte Carlo estimates of intractable path-expectations required in most cases when the diffusion model is employed.

Numerical methods developed for SDEs are usually direct extensions of corresponding schemes for ordinary differential equations. In general, they can be divided into two categories: the *strong* and the *weak* schemes. The former provide path-wise approximations of the diffusion process over all realisations of the Brownian process that drives its SDE, whereas the latter provide approximations of the diffusion probability law. Then, the closeness of a strong (resp. weak) scheme to the target diffusion is measured by the strong (resp. weak) convergence criterion. For the purposes of this review, the weak convergence criterion refers to the approximation of the distribution of the ending point X_T of the diffusion. For example, one might be interested in estimating $\mathbb{E}[g(X_T)]$ for a family of functionals $g: \mathbf{R} \mapsto \mathbf{R}$. We define explicitly both criteria in the sequel.

We will consider time discretisation schemes for the scalar diffusion (B.1) applied to the collection of time instances

$$0 = t_0 < t_1 < \dots < t_N = T ,$$

for some positive integer N . We write $\Delta_n = t_{n+1} - t_n$ and define separately the maximum step size

$$\delta = \sup\{\Delta_n; n = 0, 1, \dots, N-1\} .$$

To simplify the exposition, we assume that the time instances $t_0, \dots, t_N$ are non-random. We stress however that most of the results in the sequel would hold even if the discretisation times were random but adapted to the filtration of the Wiener process that drives (B.1). For a presentation of numerical schemes in this setting see Kloeden and Platen (1992).

The review is organised as follows. Section B.2 presents strong schemes that belong to the general family of strong Taylor approximations. Section B.3 describes some other classes of strong schemes. Section B.4 refers briefly to weak schemes

induced by the above strong ones. We finish with some general comments about the approximation methods presented here and by determining the contribution of this thesis in the context of the simulation methods for diffusion processes.

B.2 Strong Taylor approximations

The simplest and most common strong scheme for the approximation of diffusion sample paths is the *Euler-Maruyama* scheme determined by the recursion

$$Y_{t_{n+1}} = Y_{t_n} + b \Delta_n + \sigma \Delta W_n, \quad Y_0 = x_0 , \quad (\text{B.2})$$

for $0 \leq n < N$, with

$$\Delta W_n = W_{t_{n+1}} - W_{t_n} \sim \mathcal{N}(0, \Delta_n), \text{ and } b \equiv b(Y_{t_n}), \sigma \equiv \sigma(Y_{t_n}) .$$

The Euler-Maruyama scheme first appeared in Maruyama (1953, 1955). The motivation is clear; the maximum time increment δ is assumed to be sufficiently small for $b(X_t)$, $\sigma(X_t)$ to be approximately constant on each of the time intervals $[t_n, t_{n+1}]$ of increment $\Delta_n \leq \delta$.

All strong schemes presented in the sequel will be determined by a recursion of the type (B.2). In line with the notation used for the Euler-Maruyama scheme, we will omit the argument of functions appearing in the recursion when that argument is Y_{t_n} . Given a time discretisation and such a recursion one can then obtain a continuous time approximation $Y = \{Y_t; t \in [0, T]\}$ by considering, for instance, the linear interpolations between the successive discrete points. To emphasize the involvement of the maximum time increment δ at the definition of Y , we will occasionally write the approximating process as $Y = Y^\delta = \{Y_t^\delta; t \in [0, T]\}$.

According to the strong convergence criterion, the performance of a strong

scheme is measured by the absolute error

$$\mathbb{E} [|X_T - Y_T|] , \quad (\text{B.3})$$

where the joint structure of X_T, Y_T is determined by considering the same driving Wiener process for both X and Y . The approximating process $Y = Y^\delta$ converges strongly to X if (B.3) converges to 0 as $\delta \rightarrow 0$. Of more interest is the definition of the *order* of strong convergence.

Definition B.1 (Order of Strong Convergence) *We say that an approximating process $Y = Y^\delta$ converges strongly to X with order $\gamma \in (0, \infty]$ if there exist constants $K > 0$ and $\delta_0 > 0$ such that for any $\delta \in (0, \delta_0)$*

$$\mathbb{E} [|X_T - Y_T^\delta|] \leq K \delta^\gamma .$$

Under the Lipschitz and linear growth conditions which require that

$$|b(x) - b(y)| + |\sigma(x) - \sigma(y)| \leq K_1|x - y|, \quad |b(x)| + |\sigma(x)| \leq K_2(1 + |x|) ,$$

for all $x, y \in \mathbf{R}$ and some constants $K_1, K_2 > 0$, the Euler-Maruyama scheme converges strongly with order $\gamma = 0.5$, see Theorem 10.2.2 of Kloeden and Platen (1992) for a proof. Higher order schemes are also available. The *Milstein* approximation is defined as

$$Y_{t_{n+1}} = Y_{t_n} + b \Delta_n + \sigma \Delta W_n + \frac{1}{2} \sigma \sigma' \cdot \{(\Delta W_n)^2 - \Delta_n\}, \quad Y_0 = x_0 , \quad (\text{B.4})$$

where $'$ denotes derivative, and was first proposed by Milstein (1974). Under appropriate Lipschitz and linear growth conditions (see pg.350 of Kloeden and Platen, 1992) the Milstein scheme converges strongly with order $\gamma = 1.0$.

B.2.1 The general Taylor scheme

The Euler-Maruyama and Milstein schemes are particular members of the general family of strong Taylor schemes generated by truncating the stochastic Taylor expansion (also known as the Itô-Taylor expansion). In principle, one can obtain any order of strong convergence $\gamma \in \{0.5, 1.0, 1.5, \dots\}$ by considering a corresponding sufficiently large number of terms of the Itô-Taylor expansion.

The Itô-Taylor formula determines an expansion for $f(X_t)$, for any sufficiently regular $f : \mathbf{R} \mapsto \mathbf{R}$, in terms of multiple Itô's integrals and is derived by iteratively applying Itô's lemma. It was first developed by Wagner and Platen (1978). We will show the first steps of the expansion and then provide the general formula. All derivatives and stochastic integrals appearing in the sequel are assumed to exist. Applying Itô's lemma to $f(X_t)$ we find that

$$f(X_t) = f(x) + \int_0^t L^0 f(X_s) ds + \int_0^t L^1 f(X_s) dW_s , \quad (\text{B.5})$$

where we use the operators

$$L^0 = b \frac{d}{dx} + \frac{1}{2} \sigma^2 \frac{d^2}{dx^2}, \quad L^1 = \sigma \frac{d}{dx} .$$

Continuing from (B.5) and applying Itô's lemma to $L^0 f(X_s)$ and $L^1 f(X_s)$ we obtain that

$$f(X_t) = f(x) + L^0 f(x) \int_0^t ds + L^1 f(x) \int_0^t dW_s + R ,$$

for the residual

$$\begin{aligned} R &= \int_0^t \int_0^{s_2} L^{(0,0)} f(X_{s_1}) ds_1 ds_2 + \int_0^t \int_0^{s_2} L^{(1,0)} f(X_{s_1}) dW_{s_1} ds_2 \\ &+ \int_0^t \int_0^{s_2} L^{(0,1)} f(X_{s_1}) ds_1 dW_{s_2} + \int_0^t \int_0^{s_2} L^{(1,1)} f(X_{s_1}) dW_{s_1} dW_{s_2} , \end{aligned}$$

with $L^{(i,j)}f = L^i(L^j f)$ for any $i, j \in \{0, 1\}$. The expansion is further developed by applying Itô's lemma to all or some of the functionals $L^{(i,j)}f(X_{s_1})$, $i, j \in \{0, 1\}$, and then to all or some of the functionals appearing afterwards as integrands at multiple integrals, and so on.

To write down the general Itô-Taylor expansion we need to introduce some notation. Consider the set of binary vectors

$$\mathcal{M} = \{(j_1, \dots, j_l) : j_i \in \{0, 1\}, i \in \{1, \dots, l\} \text{ with } l = 1, 2, \dots\} \cup \{v\} ,$$

where v represents the trivial vector of zero length. We denote a typical element of $\mathcal{M}$ by $\alpha = (j_1, \dots, j_l)$. Also, $l(\alpha)$ will be the length of α , and $-\alpha$ the vector $\alpha \in \mathcal{M} \setminus \{v\}$ without its first element. The set $\mathcal{A} \subseteq \mathcal{M}$ is called an *hierarchical set* if

$$\mathcal{A} \neq \emptyset, \sup_{\alpha \in \mathcal{A}} l(\alpha) < \infty, \text{ and } -\alpha \in \mathcal{A} \text{ for all } \alpha \in \mathcal{A} \setminus \{v\} .$$

For any $\mathcal{A} \subseteq \mathcal{M}$ we define the set $R(\mathcal{A}) = \{\alpha \in \mathcal{M} \setminus \mathcal{A} : -\alpha \in \mathcal{A}\}$. For a given functional $g : \mathbf{R} \mapsto \mathbf{R}$, an $\alpha \in \mathcal{M}$ and some fixed time instances ρ, τ with $0 \leq \rho \leq \tau \leq T$ we define the multiple integral

$$I_\alpha [g(X_\cdot)]_{\rho, \tau} = \int_\rho^\tau \int_\rho^{s_1} \dots \int_\rho^{s_{l-1}} g(X_{s_1}) dW_{s_1}^{j_1} \dots dW_{s_{l-1}}^{j_{l-1}} dW_{s_l}^{j_l} ,$$

where for any arbitrary time index s , $dW_s^0 \equiv ds$, $dW_s^1 \equiv dW_s$. We also define the function

$$g_\alpha = L^{j_1}(\dots(L^{j_{l-1}}(L^{j_l}g))\dots) .$$

The Itô-Taylor expansion for a given hierarchical set $\mathcal{A} \subseteq \mathcal{M}$ can now be written as

$$f(X_t) = \sum_{\alpha \in \mathcal{A}} f_\alpha(x) I_\alpha [1]_{0,t} + \sum_{\alpha \in R(\mathcal{A})} I_\alpha [f_\alpha(X_\cdot)]_{0,t} .$$

From this formula one can obtain strong schemes for the approximation of the

target diffusion X by considering the identity mapping $f(x) = x$ and ignoring the multiple Itô's integrals with non-constant integrands. An arbitrary member of the family of strong Taylor approximations for X is thus determined by the recursion

$$Y_{t_{n+1}} = \sum_{\alpha \in \mathcal{A}} f_\alpha(Y_{t_n}) I_\alpha [1]_{t_n, t_{n+1}}, \quad Y_0 = x_0, \quad f(x) = x , \quad (\text{B.6})$$

for an hierarchical set $\mathcal{A} \subseteq \mathcal{M}$. For instance, the Euler-Maruyama approximation corresponds to the choice $\mathcal{A} = \{v, (0), (1)\}$ and the Milstein scheme to $\mathcal{A} = \{v, (0), (1), (1, 1)\}$.

We will now determine an explicit relationship between specific hierarchical sets and the order of convergence of the corresponding strong Taylor schemes. For an $\alpha \in \mathcal{M}$ we denote by $n(\alpha)$ the number of its elements equal to 0. Consider the subset of $\mathcal{M}$

$$\mathcal{A}_\gamma = \left\{ \alpha \in \mathcal{M} : l(\alpha) + n(\alpha) \leq 2\gamma \text{ or } l(\alpha) = n(\alpha) = \gamma + \frac{1}{2} \right\}, \quad \gamma \in \{0.5, 1.0, \dots\} ,$$

that can be trivially identified as an hierarchical set. We impose the following conditions on the functionals f_α and X , with f the identity mapping:

1. $|f_\alpha(x) - f_\alpha(y)| \leq K_1|x - y|$ for all $\alpha \in \mathcal{A}_\gamma$, $x, y \in \mathbf{R}$, for some constant $K_1 > 0$,
2. $|f_\alpha(x)| \leq K_2(1 + |x|)$ for all $\alpha \in \mathcal{A}_\gamma \cup R(\mathcal{A}_\gamma)$, $x \in \mathbf{R}$, for some constant $K_2 > 0$,
3. f_α twice continuously differentiable for all $\alpha \in \mathcal{A}_\gamma$,
4. $\{I_{\alpha-} [f_\alpha(X_\cdot)]_{0,t} ; t \in [0, \infty)\} \in \mathcal{H}_{(j_i)}$ for all $\alpha \in \mathcal{A}_\gamma \cup R(\mathcal{A}_\gamma)$,

where $\mathcal{H}_{(j)}$, $j \in \{0, 1\}$, is the set of scalar, right-continuous, with left hand limits and adapted with respect to the Wiener process filtration stochastic processes

$\{V_t; t \in [0, \infty)\}$ such that

$$\text{w.p.1, } \int_0^t |V_s|^{j+1} ds < \infty \text{ for any } t \in [0, \infty) .$$

Under conditions 1–4, the strong Taylor scheme (B.6) corresponding to the hierarchical set $\mathcal{A}_\gamma$, converges strongly to X with order γ . For a proof see Theorem 10.6.3 of Kloeden and Platen (1992).

It should be noted that it is not always possible to find explicitly the multiple integral $I_\alpha [1]_{t_n, t_{n+1}}$ for a given $\alpha \in \mathcal{M}$. However, there exist methods for its strong approximation. One such method is based on the Fourier expansion of the sample paths of the Wiener process, see ch.5 of Kloeden and Platen (1992) for more details. Still, the formulae induced by this approach are intrinsically complex and of practical use only for small $l(\alpha)$; Kloeden and Platen (1992) show the explicit formulae for the case of a Stratonovich integral and for $l(\alpha) \leq 3$.

B.3 Other families of strong schemes

In this section we show some additional strong schemes that, in cases, have some advantages compared to the Itô-Taylor approximations. For each of the approaches considered in the sequel we will not give a general formula but only present one or two examples simply to illustrate the main ideas. For more mathematical details and examples the reader is referred to Kloeden and Platen (1992).

B.3.1 Derivative free schemes

The strong Taylor schemes require the identification and the computation at specific values of several derivatives of the drift and coefficient functions. This can sometimes be computationally expensive. There exist however alternative schemes that avoid derivative calculations. In the literature they are usually referred to

as stochastic Runge-Kutta schemes in correspondence with the classical derivative free Runge-Kutta methods developed for ordinary differential equations.

A simple approach for the development of a derivative free scheme is to replace the derivatives in a given scheme with their finite difference approximations. For an example, consider the Milstein approximation (B.4). The quantity $\sigma\sigma'(Y_{t_n})$ appearing in (B.4) can be approximated by the finite difference

$$D_n = \frac{\sigma(\bar{Y}_{t_n}) - \sigma(Y_{t_n})}{\sqrt{\Delta_n}}, \quad \text{where } \bar{Y}_{t_n} = Y_{t_n} + b(Y_{t_n})\Delta_n + \sigma(Y_{t_n})\sqrt{\Delta_n},$$

yielding the numerical scheme

$$Y_{t_{n+1}} = Y_{t_n} + b\Delta_n + \sigma\Delta W_n + \frac{1}{2}D_n\{(\Delta W_n)^2 - \Delta_n\}, \quad Y_0 = x_0 .$$

It can be shown that under Lipschitz and linear growth conditions on the drift and the coefficient functions this scheme converges strongly with order $\gamma = 1.0$; for a proof see ch.11 of Kloeden and Platen (1992).

Higher order derivative free approximations are also available by appropriately modifying corresponding Taylor schemes. The expressions are becoming complex though and are of practical use only for low order schemes like the one presented here.

B.3.2 Implicit strong schemes

The basic characteristic of an implicit scheme is that the future position $Y_{t_{n+1}}$ appears in both sides of the scheme defining recursion. Incorporating additional information about the future evolution of the scheme can sometimes be proven useful. We justify this with an example in the sequel.

The implicit Euler scheme is defined by the recursion

$$Y_{t_{n+1}} = Y_{t_n} + \{\lambda b(Y_{t_{n+1}}) + (1 - \lambda)b\} \Delta_n + \sigma\Delta W_n, \quad Y_0 = x_0 ,$$

for some $\lambda \in (0, 1]$. Similarly, the implicit Milstein scheme sets

$$Y_{t_{n+1}} = Y_{t_n} + \left\{ \lambda b(Y_{t_{n+1}}) + (1 - \lambda) b \right\} \Delta_n + \sigma \Delta W_n + \frac{1}{2} \sigma \sigma' \cdot \{ (\Delta W_n)^2 - \Delta_n \} ,$$

with $Y_0 = x_0$ and $\lambda \in (0, 1]$. In both cases, the recursion determines an equation for finding $Y_{t_{n+1}}$ which typically will have to be solved using a numerical method. Under Lipschitz and linear growth conditions, the implicit Euler and Milstein schemes converge strongly with order $\gamma = 0.5$ and $\gamma = 1.0$ respectively for any choice of the tuning parameter λ . For a proof, see ch.12 of Kloeden and Platen (1992).

For simplicity, assume now that $\Delta_n = \Delta$ for all appropriate n . In some cases, a strong Taylor scheme of a given order of convergence performs dramatically differently around a critical value of the time increment Δ . In particular, there are cases when the scheme only behaves well for such small Δ 's that it is of no practical use. Implicit schemes can provide the solution to problems of this kind. For an example, consider the Ornstein-Uhlenbeck process

$$dX_t = -\beta X_t dt + dW_t, \quad t \in [0, T], \quad X_0 = x_0 ,$$

with some $\beta > 0$. Assume that $X, \bar{X}$ are the solutions of the above SDE for initial conditions $X_0 = x_0$ and $X_0 = \bar{x}_0$ respectively with $x_0 \neq \bar{x}_0$, and $Y, \bar{Y}$ the corresponding approximate processes determined by the explicit Euler scheme (B.2), all four processes driven by the same Wiener process. If the Euler scheme behaves well, then $Y_T, \bar{Y}_T$ will be close to $X_T, \bar{X}_T$ respectively. Clearly,

$$|Y_{n\Delta} - \bar{Y}_{n\Delta}| = |1 - \beta\Delta|^n |x_0 - \bar{x}_0| .$$

From the stationarity of the Ornstein-Uhlenbeck process, $|X_T - \bar{X}_T| \rightarrow 0$ as $T \rightarrow \infty$

while $|Y_T - \bar{Y}_T| \rightarrow 0$ iff

$$|1 - \beta\Delta| < 1 , \tag{B.7}$$

otherwise $|Y_T - \bar{Y}_T| \rightarrow \infty$. Thus, to obtain a good Euler approximation it is necessary that Δ satisfies (B.7) or equivalently $\Delta < 2/\beta$, a choice which for sufficiently big β can lead to a computationally inefficient scheme. On the other hand, if the approximate processes $Y, \bar{Y}$ are those determined by the implicit Euler scheme for $\lambda = 1$ then

$$|Y_{n\Delta} - \bar{Y}_{n\Delta}| = (1 + \beta\Delta)^{-n} |x_0 - \bar{x}_0| ,$$

thus the requirement (B.7) now reduces to $\Delta > 0$.

Implicit schemes are also more suitable than explicit schemes in the case of *stiff* SDEs. In brief, stiff SDEs are multidimensional SDEs where at least two of the scalar components of the solution run in a very different time scale. The consequence is that due to the component that runs slowly one has to choose a great T to reach stationarity and thus obtain a good impression of its sample paths. Then, the fast component typically forces the selection of an extremely small (comparable to the value of T chosen above) increment Δ for avoiding instabilities like the one shown in the example above. In total, one will have to run the approximation scheme over a long time period and with a very small increment, a set-up which is clearly computationally inefficient. For more details, see ch.8 of Kloeden and Platen (1992).

B.4 The weak convergence criterion

The strong schemes considered so far provide path-wise approximations of the sample paths $t \mapsto X_t$ over all realisations of the driving Brownian motion. The order of strong convergence provides a means for comparing the various schemes. We will now present the weak convergence criterion. It applies to stochastic pro-

cesses whose distribution approximates that of the objective diffusion. This time the approximating process does not need to be defined on the same probability space with the target process, and this offers greater flexibility for the development of the appropriate schemes. We will define directly the order of weak convergence according to the weak convergence criterion.

Definition B.2 (Order of Weak Convergence) *We say that an approximating process $Y = Y^\delta$ converges weakly to X with order $\beta \in \{1.0, 2.0, \dots\}$ if for any polynomial g there exist constants $K > 0$ and $\delta_0 > 0$ such that*

$$|\mathbb{E}[g(X_T)] - \mathbb{E}[g(Y_T^\delta)]| \leq K\delta^\beta,$$

for all $\delta \in (0, \delta_0)$.

Thus, the weak convergence criterion refers to the level of proximity to the marginal distribution of X_T .

All strong schemes defined earlier can be easily seen as candidate weak schemes. For example, under appropriate conditions the Euler scheme (B.2) converges weakly with order $\beta = 1.0$. The conditions are implied in the statement of the following result that connects a desired order of weak convergence with a specific truncation of the stochastic Taylor expansion.

Recall the general Taylor scheme (B.6) and the relative notation. Consider the hierarchical set of binary vectors

$$\mathcal{A}^\beta = \{\alpha \in \mathcal{M} : l(\alpha) \leq \beta\}, \quad \beta \in \{1.0, 2.0, \dots\}.$$

Let $f : \mathbf{R} \mapsto \mathbf{R}$ be the identity mapping. Under the following conditions

1. b and σ are Lipschitz continuous,
2. b and σ are $2(\beta+1)$ times continuously differentiable, and the two functions

and all their derivatives up to (and including) the $2(\beta+1)$ -th have polynomial growth,

3. $|f_\alpha(x)| \leq K(1 + |x|)$ for all $\alpha \in \mathcal{A}^\beta$, $x \in \mathbf{R}$, for some constant $K > 0$,

the Taylor scheme that uses $\mathcal{A}^\beta$ as its defining hierarchical set converges weakly with order β . This result is a special case of Theorem 14.5.1 of Kloeden and Platen (1992).

It should be noted that when a Taylor scheme is used in the weak sense, the stochastic integrals $I_\alpha[1]_{t_n, t_{n+1}}$ appearing in the scheme for some binary vector $\alpha \in \mathcal{M}$ can sometimes be replaced with simpler random variables that are easier to simulate without effect at the order of the scheme. The new variables should share similar moment properties with the stochastic integrals they are replacing. Thus, complicated methods for the strong approximation of the random variables $I_\alpha[1]_{t_n, t_{n+1}}$ required by strong Taylor schemes can be avoided, at least when $l(\alpha)$ is sufficiently small. For more details, see ch.5 of Kloeden and Platen (1992).

Weak schemes are also available for derivative free and for implicit approximations; see ch.15 of Kloeden and Platen (1992) for a short review.

B.5 Some general comments

We have only given a hint as to the vast literature related with approximate time discretisation schemes for SDEs. Nonetheless, we have presented the general Taylor scheme, for the scalar case, which is the backbone of most methods. For instance, the order of convergence of alternative schemes is typically determined by comparing them with a specific member of the family of strong Taylor approximations with known order and demonstrating that the path differences are sufficiently small for the two schemes to be of the same order.

All schemes developed in the literature are characterised by systematic error which, in principle, vanishes in a limiting sense as the increments of the time

discretisation tend to 0. Thus, a separate analysis to determine the magnitude of the bias and its sensitivity in the time increment is required each time a practitioner employs Monte Carlo estimators of diffusion expectations based on a selected time discretisation scheme. Schemes of higher order are more accurate, typically at the expense of computational efficiency.

There are even greater challenges when it comes to the development of techniques for *conditioned* diffusion processes (e.g. diffusion bridges). Traditional schemes are hard to extend to this context, and alternative strategies must be devised.

B.6 Our contribution

We will present an algorithm for the *exact* simulation of a class of diffusion processes. The Exact Algorithm (EA), presented in various versions in Articles 2, 3, 4, and 6, samples complete paths of the target SDE with no systematic error. Its simplicity is surprising, given that exact realisation of diffusion processes has been a renowned unsolved problem in the applied probability community.

The method is based on factorisations of the Wiener measure. Results in the Brownian motion literature, so far mostly motivated from theoretical considerations, are now applied to provide a solution to a practical problem. In terms of the weak and strong scheme classification in the review, EA is not a strong scheme (it is impossible to retrieve the Brownian motion driving a given sample path output of EA), but could be thought as a weak scheme of infinite order.

The intrinsic structure of EA offers great potential for further applications. We will demonstrate that in Article 4, where we take advantage of the output of EA to develop a number of methods for carrying out parametric inference for discretely observed diffusions. Further applications of EA in particle filtering (Fearnhead et al., 2006) and finance (Casella and Roberts, 2005) are currently under way.

Review C

Parametric Inference for Discretely Observed Diffusions

C.1 Introduction

Several of the recent developments in the theory of parametric estimation of discretely observed diffusion processes are reviewed. Emphasis is given on the ideas driving the methods rather than the technical details. To simplify the exposition, the methods are described as applied to scalar diffusions; the applicability to multivariate contexts is briefly discussed after the presentation of each method.

The statistical setting can be formalised as follows. The probability model is determined by the scalar SDE

$$dX_t = b(X_t; \theta) dt + \sigma(X_t; \theta) dW_t, \quad (\text{C.1})$$

driven by the standard Brownian motion $\{W_t; t \geq 0\}$. The drift $b(\cdot; \theta)$ and the coefficient function $\sigma(\cdot; \theta)$ are now assumed to be known up to the parameter vector $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbf{R}^d$. It is assumed that for any initial condition $X_0 = x_0$ and any $\theta \in \Theta$, (C.1) has a weakly unique, global solution.

Estimation of θ will be based on a single path of X observed at some discrete

non-random time instances. For simplifying the notation it is assumed that the observation instances are equidistant, thus the dataset is as follows

$$Y = Y_n = \{X_0, X_\Delta, \dots, X_{n\Delta}\}, \quad (\text{C.2})$$

for some $\Delta > 0$. Unless otherwise stated, the starting point X_0 is regarded as being fixed. Standard statistical methods do not apply in the present context since the transition density

$$p_\Delta(x, y; \theta) = \mathbb{P}[X_\Delta \in dy \mid X_0 = x; \theta] / dy,$$

is typically unavailable; notable exceptions are the geometric Brownian motion, and the Ornstein-Uhlenbeck and CIR (Cox et al., 1985) processes. Subsequently, the likelihood of the dataset (C.2)

$$\mathfrak{L}_n(\theta) = \prod_{i=1}^n p_\Delta(X_{(i-1)\Delta}, X_{i\Delta}; \theta), \quad \theta \in \Theta,$$

is usually not known analytically.

This review will consider methods built to provide accurate estimators for fixed increment Δ . Thus, when appropriate, asymptotic results demonstrating the consistency of estimators are given for $n \rightarrow \infty$ and constant Δ . For a synopsis of methods that necessitate that $\Delta = \Delta_n \rightarrow 0$ as the data size increases the reader is referred to Prakasa Rao (1999).

The structure of the review is as follows. Section C.2 considers methods that approximate the unknown likelihood function. Section C.3 presents the theory of estimating functions as applied to the diffusion setting; the unavailability of the likelihood is now dealt with by devising explicit functionals that have properties similar to the unknown score function. Section C.4 examines Bayesian MCMC methods. Section C.5 describes methods based on auxiliary models developed

in the econometrics literature. The classification of the methods follows that of the review of Sørensen (2004). Finally, we make some comments on the methods presented and determine the contribution of this thesis in the area. The Appendix shows the explicit SDEs for diffusion models mentioned in the text.

C.1.1 Preliminary notation and definitions

Some preliminary notation required in the following sections is now established. Let k, l be positive integers. For a given $f : \mathbf{R}^k \mapsto \mathbf{R}$ with arbitrary argument $x = (x_1, \dots, x_k)$, $\partial_x f$ is the vector of partial derivatives $(\partial f / \partial x_1, \dots, \partial f / \partial x_k)^\top$. For a function $f = (f_1, \dots, f_l)^\top : \mathbf{R}^k \mapsto \mathbf{R}^l$ with argument x as above, $\nabla_x f$ denotes the $l \times k$ matrix with ij -element $\partial f_i / \partial x_j$. Finally, the gaussian density is written as $\phi(u \mid \mu, \tau) = \exp\{-(u - \mu)^2 / (2\tau)\} / \sqrt{2\pi\tau}$, for $u, \mu \in \mathbf{R}$, $\tau > 0$.

The next definitions are used in the sequel. Assume that for any fixed $\theta \in \Theta$, $\sigma(\cdot; \theta)$ is positive and belongs to $C^1(\mathbf{R})$. The following process is then well defined. Let

$$\dot{X}_t = \eta(X_t; \theta), \quad \text{where} \quad \eta(u; \theta) = \int_{u_0}^u \frac{1}{\sigma(z; \theta)} dz, \quad u \in \mathbf{R}, \quad (\text{C.3})$$

and u_0 an arbitrary real number. From Itô's lemma, $\dot{X}$ solves the SDE

$$d\dot{X}_t = \dot{b}(\dot{X}_t; \theta) dt + dW_t,$$

where

$$\dot{b}(\cdot; \theta) = \frac{b(\eta^{-1}(\cdot; \theta); \theta)}{\sigma(\eta^{-1}(\cdot; \theta); \theta)} - \sigma'(\eta^{-1}(\cdot; \theta); \theta) / 2;$$

$\eta^{-1}(\cdot; \theta)$ is the inverse of $\eta(\cdot; \theta)$ and $\sigma'(\cdot; \theta)$ the derivative of $\sigma(\cdot; \theta)$. Note that (C.3) determines the unique transformation of X yielding a diffusion of unit coefficient.

Let L_θ be the generator of X , that is the operator

$$L_\theta = b(\cdot; \theta) \frac{d}{du} + \frac{1}{2} \sigma^2(\cdot; \theta) \frac{d^2}{du^2}$$

on $C^2(\mathbf{R})$. The generator $\dot{L}_\theta$ of $\dot{X}$, and any powers $L_\theta^k, \dot{L}_\theta^k, k \geq 1$, are defined in the obvious way. For a given starting point $X_0 = x$ and times $0 < t_1 < \dots < t_K$ the Euler approximation X^E of X on $\{t_1, \dots, t_K\}$ will refer to the collection of variables $\{X_{t_1}^E, \dots, X_{t_K}^E\}$ determined by the recursion

$$X_{t_k}^E = X_{t_{k-1}}^E + b(X_{t_{k-1}}^E; \theta)(t_k - t_{k-1}) + \sigma(X_{t_{k-1}}^E; \theta)W_k \sqrt{t_k - t_{k-1}}, \quad (\text{C.4})$$

for $1 \leq k \leq K$, with $t_0 = 0$, $X_0^E = x$, and $(W_1, \dots, W_K)$ a random vector from the K -dimensional standard gaussian distribution.

C.2 Likelihood approximations

The transition density $p_\Delta(x, y; \theta)$ of the diffusion X can be approximated with Monte Carlo or analytical methods. For arbitrary x, y, Δ, θ , a sequence $p_\Delta^{(K)}(x, y; \theta)$ is obtained that converges, in a sense to be made precise in the sequel, to $p_\Delta(x, y; \theta)$ for $K \rightarrow \infty$. Inference is then based on the corresponding approximation of the likelihood function

$$\mathcal{L}_n^{(K)}(\theta) = \prod_{i=1}^n p_\Delta^{(K)}(X_{(i-1)\Delta}, X_{i\Delta}; \theta), \quad \theta \in \Theta.$$

C.2.1 A closed form approximation

Ait-Sahalia (2002) provides a deterministic explicit approximation of $p_\Delta(x, y; \theta)$ by employing a Hermite expansion of it around the standard gaussian density. The expansion is called Gram-Charlier series in the literature, see for instance ch.8 of Kendall and Stuart (1977). It is known to converge if and only if the density to be expanded has tails lighter than $\exp(-y^2/4)$, see the above reference. To meet this condition and then to obtain an expansion that behaves well for $\Delta \rightarrow 0$, Ait-Sahalia (2002) carries out two transformations of the given diffusion.

Consider the diffusion X started at x on the time interval $[0, \Delta]$. To adjust

the tails of $p_\Delta(x, y; \theta)$, Ait-Sahalia (2002) transforms X to the unit coefficient diffusion $\dot{X}_t = \eta(X_t; \theta)$ defined in (C.3); let $\dot{x} = \eta(x; \theta)$ be its starting point. Then, to obtain a density which for $\Delta \rightarrow 0$ converges to the first term of the Gram-Charlier series (i.e. the standard gaussian density) and thus to prompt a rapidly convergent expansion for small Δ he employs a second transformation $\{\dot{Z}_t; t \in [0, \Delta]\}$ defined as

$$\dot{Z}_t = (\dot{X}_t - \dot{x}) \Delta^{-1/2}, \quad t \in [0, \Delta].$$

Note that $\dot{Z}_0 = 0$. Consider now the transition density

$$p_{\Delta, \dot{x}}(u; \theta) = \mathbf{P}[\dot{Z}_\Delta \in du] / du, \quad u \in \mathbf{R}.$$

A standard change-of-variable argument shows that

$$p_\Delta(x, y; \theta) = \sigma(y; \theta)^{-1} \Delta^{-1/2} p_{\Delta, \dot{x}}(\{\eta(y; \theta) - \eta(x; \theta)\} \Delta^{-1/2}; \theta). \quad (\text{C.5})$$

The Gram-Charlier expansion will apply to the standardized density $p_{\Delta, \dot{x}}(u; \theta)$; a corresponding approximation for $p_\Delta(x, y; \theta)$ will then be obtained from (C.5). The expansion develops as follows.

Consider the Hermite polynomials

$$H_j(u) = \exp\{u^2/2\} \times \frac{d^j}{du^j} \exp\{-u^2/2\}, \quad u \in \mathbf{R}, \quad j \geq 0.$$

The J -truncated Gram-Charlier series for $p_{\Delta, \dot{x}}(\cdot; \theta)$ is equal to

$$p_{\Delta, \dot{x}}^J(u; \theta) = \phi(u) \sum_{j=0}^J \lambda_j(\theta, \dot{x}, \Delta) H_j(u), \quad u \in \mathbf{R}, \quad (\text{C.6})$$

where, using the orthogonality of the Hermite polynomials ($\int_{\mathbf{R}} H_i(u) H_j(u) \phi(u) du$

is 0 for $i \neq j$ and 1 otherwise),

$$\lambda_j(\theta, \dot{x}, \Delta) = \int_{\mathbf{R}} H_j(u) p_{\Delta, \dot{x}}(u; \theta) du \equiv \mathbb{E}_\theta [H_j((\dot{X}_\Delta - \dot{x}) \Delta^{-1/2})] . \quad (\text{C.7})$$

Let $p_\Delta^J(x, y; \theta)$ denote the corresponding truncated expansion of $p_\Delta(x, y; \theta)$ from (C.5). Aït-Sahalia (2002) shows that there exists $T \in [0, \infty]$ such that for any $\Delta < T$, $p_\Delta^J(x, y; \theta) \rightarrow p_\Delta(x, y; \theta)$ as $J \rightarrow \infty$, uniformly in θ over open bounded subsets of Θ for all $(x, y) \in \mathbf{R}^2$. Also, if $\Delta < T$, the maximisers of the approximate likelihood converge a.s. to the MLE. The constant T is identified through the requirement for a gaussian bound on $u \mapsto \exp\{-(u - \dot{x})^2/(2\Delta) + \int^u \dot{b}(z) dz\}$. For several diffusions $T = \infty$.

The expectation in (C.7) is typically not analytically available thus the coefficients λ_j of the Hermite series must be approximated. For this cause, Aït-Sahalia (2002) employs the Itô-Taylor expansion, as applied to diffusion expectations, which determines that for an arbitrary polynomial f

$$\mathbb{E}_\theta [f(\dot{X}_\Delta)] = f(\dot{x}) + \sum_{k=1}^K \dot{L}_\theta^k f(\dot{x}) \frac{\Delta^k}{k!} + O(\Delta^{K+1}) .$$

He approximates λ_j with a known λ_j^K with error $O(\Delta^{K+1})$ by applying the above formula for $f(\cdot) = H_j((\cdot - \dot{x})\Delta^{-1/2})$. An explicit expression, say $p_{\Delta, \dot{x}}^{(K)}(u; \theta)$, can then be obtained for the sum in (C.6) for $J = \infty$ and λ_j replaced by λ_j^K , $j \geq 1$. The corresponding approximation $p_\Delta^{(K)}(x, y; \theta)$ of $p_\Delta(x, y; \theta)$ is

$$p_\Delta^{(K)}(x, y; \theta) = \phi(\eta(y; \theta) | \eta(x; \theta), \Delta) \sum_{k=0}^K c_k(\theta, x, y) \frac{\Delta^k}{k!} , \quad (\text{C.8})$$

for some complex in their expression, recursively defined, but analytically tractable c_k , $k \geq 0$. See Aït-Sahalia (2002) for the analytical expressions.

C.2.2 Simulated maximum likelihood

Pedersen (1995b) and Santa-Clara (1995) devised a Monte Carlo approximation of $p_\Delta(x, y; \theta)$. The idea is to impute a number of data points between each pair of the actual data so that the Euler approximation for the transition density of the enlarged dataset is made sufficiently accurate. The method develops as follows.

For a positive integer K , the time interval $[0, \Delta]$ is partitioned into $K + 1$ subintervals of equal length $\Delta^* = \Delta/(K + 1)$. Considering the diffusion X started at x at the auxiliary instances $t_k = k\Delta^*$, $1 \leq k \leq K$, and applying Chapman-Kolmogorov's equation,

$$\begin{aligned} p_\Delta(x, y; \theta) &= \int_{\mathbf{R}^K} \left\{ \prod_{k=1}^K p_{\Delta^*}(x_{k-1}, x_k; \theta) \right\} p_{\Delta^*}(x_K, y; \theta) dx_1 \cdots dx_K \quad (\text{C.9}) \\ &= \mathbb{E}_\theta [p_{\Delta^*}(X_{t_K}, y; \theta)] , \end{aligned}$$

for $x_0 = x$. The densities $p_{\Delta^*}(z, u; \theta)$, $(z, u) \in \mathbf{R}^2$, appearing in (C.9) are now approximated by the Euler density $\phi(u | z + b(z; \theta)\Delta^*, \sigma^2(z; \theta)\Delta^*)$ which for notational convenience is written as $p_{\Delta^*}^{[0]}(z, u; \theta)$. Accordingly, we denote by $p_\Delta^{[K]}(x, y; \theta)$ the integral in (C.9) after replacing the actual Δ^* -step transition densities with their approximations. Clearly,

$$p_\Delta^{[K]}(x, y; \theta) = \mathbb{E}_\theta [p_{\Delta^*}^{[0]}(X_{t_K}^E, y; \theta)] , \quad (\text{C.10})$$

where X^E is the Euler discretisation of X on $\{t_1, \dots, t_K\}$. The Euler scheme is applied to increments $\Delta^* = \Delta/(K + 1)$, thus the corresponding approximation error will decrease with increasing K . The expectation in (C.10) is estimated with the apparent Monte Carlo average over independent realisations of $X_{t_K}^E$.

Critically, the complete mapping $\theta \mapsto p_\Delta^{[K]}(x, y; \theta)$, and subsequently the corresponding likelihood approximation $\theta \mapsto \mathfrak{L}_n^{[K]}(\theta)$, can be estimated simultaneously

for all $\theta \in \Theta$ using the single stream of random elements required for its estimation at a fixed θ ; one should only fix the same gaussian seeds required for the simulation of $X_{t_K}^E$, see (C.4), for all $\theta \in \Theta$. Convexity, continuity or other smoothing characteristics of the Monte Carlo estimator of $\theta \mapsto \mathfrak{L}_n^{[K]}(\theta)$ can guarantee a passage from the point-wise convergence implied from the Monte Carlo scheme to the convergence of the maximisers of the estimator of $\theta \mapsto \mathfrak{L}_n^{[K]}(\theta)$ to the maximiser of $\theta \mapsto \mathfrak{L}_n^{[K]}(\theta)$ itself, see e.g. Geyer and Thompson (1992).

Pedersen (1995a) shows that when $K \rightarrow \infty$ then, under some boundedness assumptions on $b(\cdot; \theta)$ and $\sigma(\cdot; \theta)$, $\int_{\mathbf{R}} |p_{\Delta}^{[K]}(x, y; \theta) - p_{\Delta}(x, y; \theta)| dy \rightarrow 0$ for any x, θ , and $\mathfrak{L}_n^{[K]}(\theta) \rightarrow \mathfrak{L}_n(\theta)$ in probability for any θ . The maximisers of $\mathfrak{L}_n^{[K]}$ converge weakly to the MLE under the stronger uniform condition which requires that $\sup_{\theta \in \Theta_0} |\mathfrak{L}_n^{[K]}(\theta) - \mathfrak{L}_n(\theta)| \rightarrow 0$ in probability for all compact sets $\Theta_0 \subseteq \Theta$ and some additional weak assumptions; see Pedersen (1995a), Theorem 1. Pedersen (1995a) verifies this type of uniform convergence for a family of scalar SDEs.

Durham and Gallant (2002) presented various improvements to the above methodology. With importance sampling in mind (C.9) can be rewritten as

$$p_{\Delta}(x, y; \theta) = \mathbb{E} \left[\frac{\prod_{k=1}^{K+1} p_{\Delta^*}(U_{k-1}, U_k; \theta)}{q(U_1, \dots, U_K)} \right], \quad (\text{C.11})$$

for $U_0 = x$, $U_{K+1} = y$, a density q of support $\mathbf{R}^K$ and $(U_1, \dots, U_K) \sim q$. Durham and Gallant (2002) propose higher order approximations of the sub-transition densities $p_{\Delta^*}(z, u; \theta)$, $(z, u) \in \mathbf{R}^2$, to reduce the systematic error. They also choose densities q that (in contrast with the original approach) do take under consideration the ending point $X_{\Delta} = y$ to obtain smaller variances for their Monte Carlo estimators. Could q be the density of $[X_{t_1}, \dots, X_{t_K} | X_{\Delta} = y]$ the variance would be 0; but an analytical expression for this density is as difficult as an expression for $p_{\Delta}(x, y; \theta)$ itself. Durham and Gallant (2002) consider densities that resemble this optimal one. In contrast with the initial approach of Pedersen (1995b)

and Santa-Clara (1995) the variance of their estimators does not increase with the number K of imputed points.

C.2.3 Comments

The closed form method has the appealing property of being deterministic. The approximation error is $O(\Delta^{K+1})$ and derives from truncating a Taylor expansion in powers of Δ . In a m -dimensional setting (where the SDE is also driven by an m -dimensional Brownian motion) extending the η -transformation (C.3) requires the existence, and explicit determination, of some $\eta = \eta(x) : \mathbf{R}^m \mapsto \mathbf{R}^m$ such that $\nabla_x \eta = \sigma^{-1}$. When such a functional can be obtained the method proceeds as in the scalar case, see Aït-Sahalia (2005). Motivated by the expression of the approximation (C.8) in the one-dimensional setting, Aït-Sahalia (2005) provides analytic approximations of the transition density even when the η -transformation is impossible by postulating an expansion in powers of Δ for the transition density and determining its coefficients by plugging the expansion into Kolmogorov's backward and forward equations. The expressions for the coefficients of the expansion are highly complex and given in terms of a Taylor expansion in $y - x$, thus reducing the approximation error will require substantial effort when Δ is not very small and the diffusion paths are sufficiently volatile.

The simulated maximum likelihood method extends to a multivariate setting in the apparent way with increasing cost in the computational time, the systematic error and the variance of the importance sampling; see Brandt and Santa-Clara (2002) for an example application. The initial algorithm without the importance sampling improvements of Durham and Gallant (2002) appears to be computationally expensive and theoretically problematic since the variance of its estimator will increase with increasing number of imputed points. This owes to the fact that the expectation (C.10) uses X_{t_K} from its marginal distribution without considering the ending point $X_{\Delta} = y$. Consequently, for big K the integrand $p_{\Delta^*}^{(0)}(X_{t_K}^E, y; \theta)$

will be with large probability very close to zero and, to compensate, with a small probability large, thus its variance will be big.

C.3 Estimating functions

Consider any statistical setting when a d -dimensional parameter θ must be estimated from some dataset Y_n of size n ; let θ_0 be the true value of θ . An estimating function (EF) is a d -dimensional function of the data and the parameters, say $G_n(Y_n, \theta)$, that provides an estimator for θ_0 via the solution $\widehat{\theta}_n$ of the equation $G_n(Y_n, \theta) = 0$. G_n is called *unbiased* if

$$\mathbb{E}_\theta [G_n(Y_n, \theta)] = 0 .$$

Loosely speaking, $G_n(Y_n, \theta)$ is typically defined as an ergodic sum over the data so that as $n \rightarrow \infty$, $G_n(Y_n, \cdot)/n$ converges to a limiting function with a root at θ_0 . Some smoothing properties of $\theta \mapsto G_n(Y_n, \theta)$ then ensure that the root $\widehat{\theta}_n$ of $G_n(Y_n, \cdot)/n$ converges to the root of the limiting function, thus establishing consistency. Asymptotic normality requires a central-limit-theorem kind of result for $G_n(Y_n, \theta)/\sqrt{n}$. For a rigorous treatment of EF asymptotics see e.g. Sørensen (1999). Not coincidentally, the above arguments with $G_n(Y_n, \theta)$ replaced by the score function consist the typical steps for demonstrating the asymptotic properties of the MLE. In fact, an EF is most times defined in a way to resemble the structure of the score function.

For a complete treatment of the general theory of EFs see Heyde (1997). For the case of the diffusion setting (C.1), (C.2), and θ_0 the corresponding true parameter value, all EFs proposed in the literature try to mimic the unknown score function and thus are of the general form

$$G_n(Y_n, \theta) = \sum_{i=1}^n g(X_{(i-1)\Delta}, X_{i\Delta}, \theta) ,$$

for a $g : \mathbf{R}^2 \times \Theta \mapsto \mathbf{R}^d$.

C.3.1 Martingale estimating functions

The martingale EFs are unbiased. The structure function g is chosen in a way that

$$\mathbb{E}_\theta [g(X_{(i-1)\Delta}, X_{i\Delta}, \theta)] = 0 . \quad (\text{C.12})$$

Analytically, g is determined as the following sum

$$g(x, y, \theta) = \sum_{j=1}^J a_j(x, \theta) h_j(x, y, \theta) , \quad (\text{C.13})$$

for some positive integer J , and functions $a_j : \mathbf{R} \times \Theta \mapsto \mathbf{R}^d$, $h_j : \mathbf{R}^2 \times \Theta \mapsto \mathbf{R}$, with the condition that for any j

$$\mathbb{E}_\theta [h_j(X_0, X_\Delta, \theta) | X_0 = x] = 0 \text{ for all } x \in \mathbf{R} . \quad (\text{C.14})$$

The unbiasedness requirement (C.12) follows from (C.14) and the time homogeneity of X . The above construction yields martingale EFs; given the filtration $\mathcal{F}_i := \sigma(Y_i)$, for $Y_i = \{X_0, X_\Delta, \dots, X_{i\Delta}\}$, $1 \leq i \leq n$, it is easy to check that $\mathbb{E}_\theta [G_i(Y_i, \theta) | \mathcal{F}_{i-1}] = G_{i-1}(Y_{i-1}, \theta)$. Assuming that X is ergodic and using the martingale central limit theorem (Billingsley, 1961) the consistency and asymptotic normality of $\widehat{\theta}_n$ follow under weak conditions, see Sørensen (1999); Bibby and Sørensen (1995).

In almost all cases in the literature functions h_j with the property (C.14) are obtained as

$$h_j(x, y, \theta) = f_j(y, \theta) - \mathbb{E}_\theta [f_j(X_\Delta, \theta) | X_0 = x] , \quad (\text{C.15})$$

for some $f_j : \mathbf{R} \times \Theta \mapsto \mathbf{R}$. The functions f_j , $1 \leq j \leq J$, are called the *base* of the EF; the a_j appearing in (C.13) are called *weights*. There are no specific rules

for the choice of J and of the base functions f_j . In applications of EFs in the literature two basic approaches have been followed.

The first involves polynomial EFs with base $f_j(y, \theta) = y^{k_j}$, for some positive integers k_j . Linear EFs ($J = 1, k_1 = 1$) were used in Bibby and Sørensen (1995); quadratic EFs ($J = 2, k_1 = 1, k_2 = 2$) in Bibby and Sørensen (1996, 1997). For these base functions the conditional expectation term in (C.15) is analytically available only for a few specific SDEs (see the above references). Generally, in any case when the choice of a base function results in an intractable expectation, the latter is numerically estimated with the apparent Monte Carlo procedure, see e.g. Bibby and Sørensen (1995); Kessler and Paredes (2002). The family of EFs that are numerically determined are termed *non-explicit* EFs in the literature in contrast with the *explicit* EFs that do not require simulations.

Kessler and Paredes (2002) study the effect of replacing the expectation in (C.15) with its Monte Carlo estimation that averages over a fixed number of independent draws, say M , from $[X_\Delta | X_0 = x]$. Assume that this distribution is approximated under the Euler scheme (C.4) for a time discretisation of increment $\delta = \delta_n$. Under some regularity conditions, if $\delta_n = o(n^{-1/2})$ then the EF estimator is still consistent and asymptotically normal and its asymptotic variance is only $(1 + 1/M)$ times bigger the asymptotic variance of the estimator from the EF that calculates explicitly the expectations. See Kessler and Paredes (2002) for some more general results.

The second family of martingale EFs builds on the eigenfunctions of the generator L_θ of the diffusion; the motivation has been precisely the determination of base functions with known conditional expectations. An $f : \mathbf{R} \times \Theta \mapsto \mathbf{R}$ is an eigenfunction of L_θ with eigenvalue the real λ_θ if

$$L_\theta f(y, \theta) = -\lambda_\theta f(y, \theta) . \quad (\text{C.16})$$

Under weak conditions, (C.16) implies that

$$\mathbb{E}_\theta [f(X_\Delta, \theta) | X_0 = x] = \exp(-\lambda_\theta \Delta) f(x, \theta) ,$$

see Kessler and Sørensen (1999). Thus, if $f_j, 1 \leq j \leq J$, are distinct eigenfunctions with corresponding eigenvalues $\lambda_{j,\theta}$ then a martingale EF is obtained by specifying

$$h_j(x, y, \theta) = f_j(y, \theta) - \exp(-\lambda_{j,\theta} \Delta) f_j(x, \theta) .$$

This construction of the EF is possible only for a small number of SDEs for which (C.16) can be solved explicitly, see the examples in Kessler and Sørensen (1999).

Once the base functions f_j are selected with one of the two approaches above or any other method, any choice of the weights a_j (given that some regularity properties are satisfied, see Sørensen, 1999) will result in an estimator with the appropriate asymptotic properties. The *optimal* weights a_j^* , in the sense that the corresponding EF is closest to the score function, can be found using the general theory of optimal EFs, see for example Godambe and Heyde (1987). From Theorem 3.1 of Sørensen (1997)

$$a^*(x, \theta) = \{ \mathbb{E}_\theta [\nabla_\theta h(X_0, X_\Delta, \theta) | X_0 = x] \}^\top \{ \text{Var}_\theta [h(X_0, X_\Delta, \theta) | X_0 = x] \}^{-1} ,$$

where a^* is the $d \times J$ matrix with columns a_j^* , while $h = (h_1, \dots, h_J)^\top$. In most cases a^* is not analytically available and some approximation is used instead, see e.g. Bibby and Sørensen (1995).

C.3.2 Other families of EFs

Assume that X is ergodic with stationary density $\pi(\cdot; \theta)$. *Simple* EFs are determined as

$$G_n(Y_n, \theta) = \sum_{i=0}^n g(X_{i\Delta}, \theta) ,$$

for a $g = (g_1, \dots, g_d)^\top : \mathbf{R} \times \Theta \mapsto \mathbf{R}^d$ such that

$$\int_{\mathbf{R}} g_j(y, \theta) \pi(y; \theta) dy = 0, \quad 1 \leq j \leq d. \quad (\text{C.17})$$

The family of simple EFs was studied in Kessler (2000) who notes that they can infer only about the parameters appearing at $\pi(\cdot; \theta)$ and gives conditions on g , X so that the corresponding estimator is consistent and asymptotically normal. Kessler (2000) also determines the optimal simple EF which is in practice intractable except for trivial cases.

The motivation for simple EFs is the availability of a large family of functionals g that satisfy (C.17). Kessler (2000) considers the choice $g(y, \theta) = \partial_\theta \log \pi(y; \theta)$, since the stationary density is typically available for scalar diffusions, and also $g(y, \theta) = (L_\theta f_1(y, \theta), \dots, L_\theta f_d(y, \theta))^\top$ for some $f_j : \mathbf{R} \times \Theta \mapsto \mathbf{R}$, $1 \leq j \leq d$. See also Bibby and Sørensen (2001) for an application where simple EFs are supplemented by martingale EFs to infer about the complete parameter vector.

C.3.3 Comments

An important property of the EF method is that its estimators are by definition less accurate, in terms of their asymptotic variance, than the MLE. Kessler (2000) describes a case when the optimal simple EF cannot infer about the parameter except for huge data size. Simple EFs miss all the information from the conditional structure of the data. Martingale EFs based on the eigenfunctions of the generator L_θ are tailored on the specific structure of the diffusion and are expected to behave better than polynomial EFs; the former though are rarely available. In a multi-dimensional context, martingale EFs are obtained in the same way as for the scalar case, though conditional expectations are now harder to obtain so the EFs are usually non-explicit. Also, the eigenfunctions of L_θ are now almost impossible to find, see section 3 of Bibby et al. (2002). In the same paper an application to

a stochastic volatility model is described.

C.4 Bayesian-MCMC Methods

Assume that θ is given a prior density $\pi(\theta)$. The likelihood of the dataset Y in (C.2) is intractable and so is the posterior density $\pi(\theta | Y)$ on which inference in a Bayesian context is based. The suggested MCMC methods for overcoming the unavailability of $\pi(\theta | Y)$ involve augmenting Y with the missing paths between consecutive data points. A Gibbs sampler is then applied to the joint posterior density of θ and the augmented data. More accurately, the augmented data will be a finite projection of the continuous time missing paths or a transformation of them. These two approaches are now described in more detail.

C.4.1 High frequency augmentation (HFA)

The title is borrowed from Jones (1998). The method can be thought as a transfer of the simulated maximum likelihood approach of Section C.2.2 into a Bayesian context.

For each pair of consecutive data points $X_{(i-1)\Delta}, X_{i\Delta}$ consider the K latent elements $X_{i,k} = X_{t_{ik}}$ with $t_{ik} = (i-1)\Delta + k\Delta/(K+1)$ where $1 \leq k \leq K$. Let

$$Z := \{Z_i; 1 \leq i \leq n\}, \quad \text{with } Z_i := \{X_{i,k}; 1 \leq k \leq K\}.$$

The complete dataset Y, Z involves instances of the diffusion process of time increment $\Delta^* = \Delta/(K+1)$. Similarly to Section C.2.2, K can be chosen big enough so that the Euler approximation (C.4) when applied to Y, Z is sufficiently accurate. Thus, the likelihood of Y, Z can be considered as being approximately equal to

$$\pi^K(Y, Z | \theta) = \prod_{i=1}^n \left\{ \prod_{k=1}^{K+1} \phi(X_{i,k} | X_{i,k-1} + b(X_{i,k-1}; \theta)\Delta^*, \sigma^2(X_{i,k-1}; \theta)\Delta^*) \right\},$$

where $X_{i,0} \equiv X_{(i-1)\Delta}$, $X_{i,K+1} \equiv X_{i\Delta}$. A Gibbs sampler will now apply to

$$\pi^K(\theta, Z | Y) \propto \pi^K(Y, Z | \theta) \pi(\theta),$$

initiated with some $\theta^{(0)}, Z^{(0)}$ and proceeding by updating $\theta^{(j)}, Z^{(j)}$ to $\theta^{(j+1)}, Z^{(j+1)}$ with

1. $\theta^{(j+1)}$ the output of a Gibbs or M-H step on $\pi^K(\theta | Z^{(j)}, Y)$,
2. $Z_i^{(j+1)}$ the output of a Gibbs or M-H step on $\pi^K(Z_i | \theta^{(j+1)}, Y)$, $1 \leq i \leq n$,

where M-H stands for the Metropolis-Hastings MCMC algorithm. For details on the Gibbs and M-H methods see e.g. Gelfand and Smith (1990). Step 2 exploits the conditional independence of $\{Z_i; 1 \leq i \leq n\}$ given Y, θ . The density appearing in this step is proportional to the i -th factor of $\pi^K(Y, Z | \theta)$. The MCMC protocol implies that for $j \rightarrow \infty$, the distribution of $\theta^{(j)}, Z^{(j)}$ converges to $\pi^K(\theta, Z | Y)$. Inference about the posterior density of θ will be based on the marginal trajectory of $\theta^{(j)}$.

The theory developed so far presents the common framework of the MCMC algorithms for scalar diffusions introduced in Jones (1998); Eraker (2001); Elerian et al. (2001). Each of these algorithms provides a specific method for carrying out the two steps of the MCMC algorithm described above. For the update of θ they all suggest a model-specific consideration of the full conditional density of θ . For the Z_i 's, Jones (1998) and Eraker (2001) propose a cyclic M-H algorithm that updates each univariate element of Z_i conditionally on the current values of its adjacent elements. Elerian et al. (2001) notice that the mixing of the MCMC algorithm improves when Z_i is divided into blocks of random size and a M-H algorithm is applied iteratively to each of these blocks conditionally on the current value of the rest.

C.4.2 Missing path transformation

Roberts and Stramer (2001) noticed that when the diffusion coefficient σ is a function of θ the rate of convergence of the trajectory $\theta^{(j)}, Z^{(j)}$ to the target distribution $\pi^K(\theta, Z | Y)$ can be made arbitrary slow for increasing number of imputed data. The degeneracy of the MCMC algorithm for $K \rightarrow \infty$ can be avoided by appropriately transforming the continuous time missing paths. Consider the missing paths $\{X_{i,t}; t \in [0, \Delta]\}$, $1 \leq i \leq n$, determined as

$$X_{i,t} = X_{(i-1)\Delta+t}, \quad t \in [0, \Delta].$$

Similarly to Section C.2.1, Roberts and Stramer (2001) consider the η transformation given in (C.3). Recall that $\eta(X; \theta)$ solves an SDE of unit diffusion coefficient. Consider the following transformation of the missing paths

$$\check{X}_{i,t} := \eta(X_{i,t}; \theta) - \frac{\eta(X_{(i-1)\Delta}; \theta) (\Delta - t) + \eta(X_{i\Delta}; \theta) t}{\Delta}, \quad t \in [0, \Delta]. \quad (\text{C.18})$$

Let $\check{X}_i = \{\check{X}_{i,s}; s \in [0, \Delta]\}$ and $\check{X} = \{\check{X}_i; 1 \leq i \leq n\}$. The choice of the above transformation is not accidental. Conditionally on θ, Y each $\check{X}_i$ is absolutely continuous w.r.t. a Brownian bridge of length Δ starting and finishing at 0. Thus, it is now possible to find the joint density of $[Y, \check{X} | \theta]$ under a dominating measure that does not depend on θ , namely the product of the n -dimensional Lebesgue measure with n independent standard Brownian bridge laws. A Gibbs sampler will then apply to the posterior density $\pi(\theta, \check{X} | Y)$, which can be found analytically. The transformed missing paths $\check{X}_i$, $1 \leq i \leq n$, are updated using an Independence sampler step with Brownian or Ornstein-Uhlenbeck bridge proposals, see Roberts and Stramer (2001) for details. Riemann and stochastic integrals appearing in the application of the Gibbs sampler are approximated in the obvious way by considering a finite projection of some size K of $\check{X}_i$, $1 \leq i \leq n$. The MCMC

algorithm does not degenerate as $K \rightarrow \infty$.

C.4.3 Comments

The rate of convergence of the actual, i.e. without the involvement of M-H steps, Gibbs sampler in the above algorithms, is equal to

$$\sup \{ \text{corr} [f(\theta), g(\text{augmented data}) | Y] \} ,$$

the supremum applied over all real-valued functionals f, g ; see Amit (1991). The transformation (C.18) reduces the information about θ contained in the augmented data, so the path-transformation algorithm converges faster than the HFA ones. Simultaneously, the former algorithm overcomes the paradox of a deteriorating MCMC algorithm for increasing accuracy at the computation of the posterior density $\pi^K(\theta, Z | Y)$ that characterises the HFA approach.

The HFA algorithms extend easier to multivariate settings and stochastic volatility models, though with multiplicative cost at their convergence properties. For the case of the path-transformation algorithm, modifying a multidimensional SDE into an SDE of unit diffusion coefficient requires a solution to a multidimensional differential equation; see the comments for the closed form likelihood approximation method of Section C.2.1 where the same modification is needed. All algorithms carry discretisation bias due to the finite approximation of continuous time paths.

C.5 Auxiliary model methods

This section considers inference methods developed in the econometric literature. In particular, the Indirect Inference theory, introduced by Smith (1993) and further developed by Gouriéroux et al. (1993), and the Efficient Method of Moments (EMM), introduced by Gallant and Tauchen (1996), are presented. Both methods

can be applied to general probability models. However, in the sequel we have adopted a diffusion-tailored description of the two approaches.

Both methods require simulations from the base model and the introduction of an auxiliary model with known transition density. The auxiliary model is typically obtained in terms of an alternative transition density $q_\Delta(x, y; \rho)$ with parameter $\rho \in R \subseteq \mathbf{R}^r$ where $r \geq d$ (e.g. the density of the Euler approximation). We will need the quasi-MLE

$$\rho_n := \underset{\rho \in R}{\operatorname{argmax}} \sum_{i=1}^n \log q_\Delta(X_{(i-1)\Delta}, X_{i\Delta}; \rho) .$$

Having selected the auxiliary model, Indirect Inference and EMM will then develop in three steps. Recall that θ_0 denotes the true parameter value.

Step 1: Construction of a r -dimensional function of the quasi-MLE and the parameter, say $F(\rho_n, \theta)$, determined by the auxiliary model. The explicit function will be different for each of the two approaches. When $n \rightarrow \infty$, $F(\rho_n, \theta) \rightarrow F_\infty(\theta)$ uniformly in θ (weakly or strongly) for some limiting function $\theta \mapsto F_\infty(\theta)$ with $F_\infty(\theta_0) = 0$.

Step 2: F will not be analytically available. Instead, an estimate of it, say $\theta \mapsto F_k(\rho_n, \theta)$, will be obtained using the *simulated* datasets

$$Y_k^\theta = \{X_0^\theta, X_\Delta^\theta, \dots, X_{k\Delta}^\theta\}, \quad \theta \in \Theta, \quad (\text{C.19})$$

each of size $k \geq 1$. For a fixed θ , Y_k^θ corresponds to a simulation of the base model (C.1) for parameter value θ at the time instances $0, \Delta, \dots, k\Delta$. The simulation is carried out using a single stream of random elements for all $\theta \in \Theta$, typically by employing the Euler scheme (C.4) on some equidistant time instances of some small increment (small enough so that the simulation is presumed to be from the actual model) and fixing the same gaussian seeds for all $\theta \in \Theta$.

Step 3: An estimator is obtained as

$$\theta_{n,k} = \operatorname{argmin}_{\theta \in \Theta} F_k(\rho_n, \theta)^\top \Omega F_k(\rho_n, \theta) , \quad (\text{C.20})$$

and conditions are specified for F_k and the diffusion model for the estimator to be consistent and asymptotically normal. The $r \times r$ positive definite weight matrix Ω is chosen to minimise the asymptotic variance of the estimator.

In the sequel the objective function $F(\rho_n, \theta)$ and its estimation $F_k(\rho_n, \theta)$ are defined for each of the two methods and additional details required for understanding the methodology are given. Some convergence results are given merely for demonstrating the main ideas, thus we make no reference to regularity conditions under which these results hold and to the type (strong or weak) of convergence. Both methods require that X be ergodic. Recall that $\pi(\cdot; \theta)$ denotes the stationary density for X .

C.5.1 Indirect Inference

Intuitively, the Indirect Inference estimator is the parameter value in Θ for which the simulated data (C.19) are closest to the real data. Discrepancy between datasets is determined from their effect to the MLE of the auxiliary model.

The method is based on the notion of the *binding function* which we now explain. Assume that the quasi-MLE corresponding to the simulated data

$$\rho_k^\theta = \operatorname{argmax}_{\rho \in R} \sum_{i=1}^k \log q_\Delta(X_{(i-1)\Delta}^\theta, X_{i\Delta}^\theta; \theta), \quad \theta \in \Theta , \quad (\text{C.21})$$

converges as $k \rightarrow \infty$ to some $\rho^\theta \in R$. Intuitively, for a given $\theta \in \Theta$, ρ^θ identifies the pseudo-true process ultimately suggested by the inference procedure under the (misspecified) auxiliary model when the data are actually coming from the structure model with parameter θ . The binding function is the mapping $\theta \mapsto \rho^\theta$;

for a formal definition see Gourieroux et al. (1993).

The objective function is now defined as $F(\rho_n, \theta) = \rho^\theta - \rho_n$ and (under conditions) for $n \rightarrow \infty$ converges uniformly in θ to $F_\infty(\theta) = \rho^\theta - \rho^{\theta_0}$. The binding function is typically analytically intractable and so is $F(\rho_n, \theta)$. A simple estimate of $F(\rho_n, \theta)$ is determined as

$$F_k(\rho_n, \theta) = \rho_k^\theta - \rho_n, \quad \theta \in \Theta .$$

Gourieroux et al. (1993) study the asymptotic properties of the corresponding Indirect Inference estimator $\theta_{n,k}$ in (C.20) when $k = k(n) = nK$ for a fixed $K \geq 1$, provide conditions for its consistency and find the weight matrix Ω that minimises its asymptotic variance.

C.5.2 Efficient Method of Moments (EMM)

The “efficient” term at the name of the method emphasizes the property of its estimators to be asymptotically equivalent to the MLE when the auxiliary model encompasses the actual model (a property also shared by the Indirect Inference estimator). It also highlights applications of EMM in the literature in conjunction with the specification of a *sequence* of auxiliary models that yield in the limit estimators asymptotically equivalent to the MLE; see e.g. Gallant and Long (1997).

Let $s_\Delta(x, y; \rho) = \partial_\rho \log q_\Delta(x, y; \rho)$. Consider the quasi-score function

$$m_k^\theta(\rho) = \frac{1}{k} \sum_{i=1}^k s_\Delta(X_{(i-1)\Delta}^\theta, X_{i\Delta}^\theta; \rho), \quad \rho \in R ,$$

as applied to the simulated datasets, and assume that for $k \rightarrow \infty$ it converges uniformly in ρ to

$$m^\theta(\rho) = \int_{\mathbf{R}^2} s_\Delta(x, y; \rho) \pi(x; \theta) p_\Delta(x, y; \theta) dx dy, \quad \rho \in R ,$$

so that the pseudo-true parameter ρ^θ implied by carrying out inference based on the auxiliary model is uniquely specified from the equation $m^\theta(\rho^\theta) = 0$. The objective function is defined as $F(\rho_n, \theta) = m^\theta(\rho_n)$ with limiting function $F_\infty(\theta) = m^\theta(\rho^{\theta_0})$. It is estimated as

$$F_k(\rho_n, \theta) = m_k^\theta(\rho_n) .$$

The asymptotic properties of the corresponding estimator defined in (C.20) are described for instance by Gallant and Long (1997).

C.5.3 Comments

A necessary condition for implementing Indirect Inference or EMM is that the structure model can be simulated; otherwise, both methods apply under fairly weak conditions. Thus, they can be easily adapted to multivariate diffusions and stochastic volatility models. The choice of the auxiliary model is critical since it effects directly the finite sample and asymptotic properties of the estimators. The estimators improve as the auxiliary model approximates the structure model; when the structure model is nested within the auxiliary model the estimators are asymptotically equivalent to the MLE. A convenient characteristic of both methods is that they provide a simple test for identifying model misspecification. Analytically, under the null hypothesis of correctness of the structure model, the statistic

$$n F_k(\rho_n, \theta_{n,k})^\top \Omega_n F_k(\rho_n, \theta_{n,k}) ,$$

where Ω_n is a specific weight matrix determined in a different way for each of the two methods, is asymptotically distributed (for $n, k \rightarrow \infty$) as a chi-square with $r - d$ degrees of freedom; for more details see Gouriéroux et al. (1993) and Gallant and Long (1997).

Comparing the two methods, one should note that they are equivalent in terms of the asymptotic variance of the resulted estimators for a certain correspondence

of the weight matrix used in each case (see Gouriéroux et al., 1993). From a computational point of view however, EMM is superior to Indirect Inference since for the latter each evaluation of the quadratic form in (C.20) at some θ requires an optimisation algorithm for finding the quasi-MLE in (C.21). Subsequently, most current econometric applications base their inference objectives on EMM.

Critically, and similarly to the case of EFs in Section C.3, Indirect Inference and EMM yield estimators inferior to the MLE in terms of asymptotic variance. Even in cases when theory suggests asymptotic proximity to the MLE, the finite sample properties of the estimators of the two methods can in cases be severely dissatisfactory; see e.g. the numerical studies by Zhou (2001) and Duffee and Stanton (2004).

C.6 Some general comments

All methods introduced so far in the literature for carrying out parametric inference for discretely observed diffusions are computationally demanding. The number of researchers currently working in the area and the amount of the related published work demonstrate that a definite solution to the unidentifiability of the diffusion transition density has yet to appear. Such a conclusion is even more evident in the case of multi-dimensional or latent diffusion models.

A comparison of the existing methods is not straightforward and has not been carried out in this thesis. For the most relevant work in this direction the reader is referred to Zhou (2001); Jensen and Poulsen (2002). In general, likelihood-based methods seem more natural than alternative approaches (EFs, EMM or Indirect Inference) since the latter are not guaranteed to retrieve as much information from the data as necessary to carry out reliable inference.

Methods based on simulations typically incorporate systematic error due to the introduction of time discretisation schemes for the approximation of the diffusion

model. Sometimes a big number of intermediate data points must be imputed for the approximation to be close to the actual model, with consequences to the computational cost for implementing the corresponding algorithm.

C.7 Our contribution

In Article 4, we will introduce a number of new methods for carrying out inference for a class of discretely observed diffusion processes. We present an MCMC, a Monte-Carlo EM, and a likelihood estimation algorithm. All methods build on the structure of the Exact Algorithm (EA). Our algorithms avoid any systematic bias, and involve only Monte Carlo error. They appear to be computationally efficient; only a small (random) number of the missing data locations need to be imputed to eliminate the discretisation bias. Article 4 has been accepted as a discussion paper in JRSSB.

In Article 5, we will explore the likelihood estimation algorithm and study its theoretical properties. We present in detail the Brownian path couplings required for extending its applicability to a particular class of diffusion processes. We also give a general recipe for the implementation of the algorithm that does not necessitate familiarity with the underlying theory.

Part II

Articles

Article 1

One-Shot CFTP; Application to a Class of Truncated Gaussian Densities

1.1 Introduction

Although Perfect Simulation (Propp and Wilson, 1996) is now well-known and has been used in a variety of MCMC contexts, it has proved difficult to apply effectively in high-dimensional situations without specific symmetry to guide the algorithm construction. Thus in particular MCMC problems from Bayesian statistics are difficult for Perfect Simulation (see Murdoch and Green, 1998a, for a first attempt). Furthermore, continuous state space problems are intrinsically difficult for any technique based on coupling, since intricate constructions are needed to ensure coupling at all, let alone allow the construction of efficient coalescence rules for complex Markov chains.

In this paper we shall bring together and extend a collection of generic techniques for explicit coupling of continuous state space Markov chains. We will show how these methods can be applied effectively for high-dimensional Gibbs samplers

and related algorithms. We illustrate our methods through an example which will be explored in detail through the paper and involves a gaussian density restricted to an arbitrary rectangular region. However we hope to illustrate that our methods have great potential in more general situations.

Our example will achieve exact sampling from multivariate gaussian densities restricted to rectangular regions of the form:

$$\pi(\mathbf{x}) \propto \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu)^\top Q (\mathbf{x} - \mu) \right\} \cdot \mathbb{I}_S(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^d, \quad (1.1)$$

for $\mu \in \mathbf{R}^d$, $Q = (q_{ij}) \in \mathbf{R}^{d \times d}$ a positive definite, symmetric matrix and support $S = \{\mathbf{x} \in \mathbf{R}^d : \mathbf{a} \preceq \mathbf{x} \preceq \mathbf{b}\}$, $\mathbf{a}, \mathbf{b} \in \bar{\mathbf{R}}^d$. ($\preceq$) is the natural partial ordering on $\mathbf{R}^d$, $\mathbb{I}_S$ the indicator function of S , $\bar{\mathbf{R}}$ the extended real line $\mathbf{R} \cup \{-\infty, \infty\}$. We call Q the precision matrix. Considering location-scale re-parameterisation we can assume that $\mu \equiv \mathbf{0}$ and $q_{ii} \equiv 1$, $1 \leq i \leq d$.

The abbreviation TN will refer to an arbitrary member of the family of distributions with densities of the type (1.1). When a random vector $\mathbf{X}$ is distributed according to π we write $\mathbf{X} \sim TN_d(\mu, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}}$. Whenever some vector $\mathbf{u} \in \mathbf{R}^d$ is mentioned u_i denotes its i -th element and $\mathbf{u}_{-i}$ the complete vector without its i -th element, $1 \leq i \leq d$. We use bold fonts for vectors preserving the right to drop this convention whenever needed.

Our methodology deals most easily with cases when the precision matrix is *Stieltjes* (Varga, 1962, ch.3):

Definition 1.1 A matrix $V = (v_{ij}) \in \mathbf{R}^{d \times d}$ is a *Stieltjes matrix* if it is symmetric, positive definite with $v_{ij} \leq 0$ for all $i \neq j$.

A Stieltjes precision matrix induces a stochastically monotone (w.r.t. ($\preceq$)) Gibbs sampler on π and this is precisely the property that makes possible the development of the Perfect Simulation algorithms that follow.

Econometrics (Geweke, 1991; Hajivassiliou and McFadden, 1998) and signal

processing (Djuric and Kotecha, 1999; Rodriguez-Yam et al., 2002; Troughton and Godsill, 1999) literature provides several cases when TN s arise and different methods for sampling from them. The univariate case is easy to deal with using appropriate rejection sampling methods, see for example Geweke (1991); Robert (1995). Rejection sampling algorithms in multivariate cases involve comparison functions like the corresponding untruncated density, the product of univariate exponential (Hajivassiliou et al., 1996) or gaussian densities restricted to the support of the target TN and sequentially generated densities (Hajivassiliou and McFadden, 1998). All these techniques perform poorly in high dimensions or for constraint regions away from the bulk of the underlying untruncated density. Many have recommended the use of the Gibbs sampler (Djuric and Kotecha, 1999; Geweke, 1991; Hajivassiliou and McFadden, 1998; Robert, 1995; Rodriguez-Yam et al., 2002) or some other MCMC technique (Troughton and Godsill, 1999). These methods return correlated draws from a distribution that approximates (in a suitable sense) the actual target distribution. Recently, a Perfect Simulation method for TN s restricted to $[0, \infty)^d$ has been presented (Philippe and Robert, 2003) which exploits a monotonicity property induced by the MCMC technique of Slice sampler. Its algebraic difficulties render it manageable only for $d \leq 3$. It is also not exact in the sense that a vector with sufficiently big values is used instead of $\{+\infty\}^d$.

Our methodology, as applied to a TN , lies close to the Gibbs sampler approach and could be thought of as a way of turning the approximate results of the Gibbs method into exact ones. The Gibbs sampler when applied to a TN exploits the fact that the univariate full conditional distributions are still TN s. Indeed, if $\mathbf{X} \sim \pi$ then:

$$X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i} \sim TN_1(\mu_i(\mathbf{x}_{-i}), 1)_{|a_i}^{b_i}, \text{ for } \mu_i(\mathbf{x}_{-i}) = - \sum_{j:j \neq i}^d q_{ij} x_j. \quad (1.2)$$

The algorithms to be presented follow the Read-Once coupling from the past (Ro-

CFTP) protocol (Wilson, 2000b), so we begin by presenting in Section 1.2 a version of the Ro-CFTP idea called *One-Shot CFTP* and introduced in Roberts and Rosenthal (2002). In Section 1.3 we introduce the coupling technique of the *Monotone Independence Coupler* (MIC) necessary for the development of the Perfect Simulation algorithms that follow.

Two algorithms will be presented. The *case-specific* (Section 1.4) one is the first that comes in mind under the certain structure of the precision matrix and relies on the cdf Φ of the $\mathcal{N}(0, 1)$ distribution and its inverse Φ^{-1} . Values of $\Phi(x)$ and $\Phi^{-1}(u)$ for big $|x|$ and $u \downarrow 0, u \uparrow 1$ must be treated with caution because of the finite accuracy of computing systems and this sometimes restricts the efficiency of the algorithm.

The *pure* algorithm (Section 1.5) does not use at any point Φ or its quantiles. It is more complicated than the case-specific one but more general. Although strictly speaking it is not needed for the truncated gaussian example (apart from the numerical issues mentioned above which nonetheless can sometimes be critical for the applicability of the algorithm, see the last example in Section 1.7.2) we will describe it in detail and implement it with a view to demonstrating the general applicability of our methodology. Both algorithms exploit the Stieltjes property of the precision matrix; the case-specific algorithm by constructing a monotone Gibbs sampler and the pure algorithm by devising convergent upper and lower dominating processes for the Gibbs sampler Markov chain.

In Section 1.6 we show that the applicability of the above algorithms can be easily extended to a class of truncated gaussian densities wider than the one with Stieltjes precision matrix and that similar algorithms can be devised even for non-gaussian densities. In Section 1.7 we run the algorithms and present several results concerning their correctness and efficiency. It turns out that they draw efficiently from the target distribution even for high dimensions. We finish with some conclusions in Section 1.8.

1.2 One-Shot CFTP

The One-Shot coupling has been introduced in Roberts and Rosenthal (2002) in the context of obtaining computable bounds for the convergence of Markov chains. Our approach will be to use the method as an ingredient in a Ro-CFTP algorithm. The construction proceeds as follows. Suppose we have two Markov chains $(X_n)_0^\infty, (Y_n)_0^\infty$, say on $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$, governed by the same transition probability kernel but with different initial distributions which are *coupled* in a way that $|X_n - Y_n| \rightarrow 0$ as $n \rightarrow \infty$ (note that we drop the bold font representation of vectors for this section). An attempt to unite the two paths is made when their distance is small and this occasionally leads to stringent bounds on the total variation distance $\sup_{A \in \mathcal{B}(\mathbf{R}^d)} |\mathbf{P}[X_n \in A] - \mathbf{P}[Y_n \in A]|$. We adjust the One-Shot coupling idea within the context of the Ro-CFTP algorithm.

In the sequel, we adopt the Stochastic Recursive Sequence representation of a Markov chain and use *random update functions* as means of describing the dynamics of the chain and coupling its one-step transitions. Assume that we need to sample from a (possibly unnormalised) density π on $\mathbf{R}^d$ w.r.t. some reference measure. For convenience we denote by π both this density and the probability measure induced by it. Let S be the support of π .

The CFTP protocol (Propp and Wilson, 1996) involves the development of an ergodic (usually discrete-time) Markov chain with stationary distribution π and the parallel realisation of paths of the chain starting from everywhere in the support of π in a way that it is possible all paths to merge into a single trajectory in finite time. The paths are run forward with starting points taken increasingly back in time until a random moment $T < 0$ is found when they all get projected onto a single point at time 0. This single output is proved to be distributed according to π . During this repeated excursion into the past random seeds used for the realisation of the paths are stored and recalled any time paths coming from deeper into the past meet the time instances where these seeds were first used.

Unlike the basic CFTP idea, the Ro-CFTP technique does not require the storage of random seeds since it evolves forward in time. It relies on the development of a number, say k , of random update functions ϕ_i , $1 \leq i \leq k$, and their synthesis to a compound random update function which in our case will have the structure:

$$\Phi(x, \bar{U}_k) := \phi_k(\cdots(\phi_2(\phi_1(x, \bar{U}_1), \bar{U}_2), \dots, \bar{U}_k), \quad x \in S,$$

or in an abbreviated form:

$$\Phi(x) = \phi_k \circ \cdots \circ \phi_2 \circ \phi_1(x), \quad x \in S. \quad (1.3)$$

We have denoted by U_i the random elements needed for the first time at the development of ϕ_i . Also, $\bar{U}_i := (U_1, U_2, \dots, U_i)$, $1 \leq i \leq k$. The update rule Φ must *preserve* π :

$$\text{if } \mathcal{P}(x, dy) := \mathbf{P}[\Phi(x) \in dy], \quad \text{then } \int_S \pi(dx) \mathcal{P}(x, dy) = \pi(dy), \quad (1.4)$$

or in a less troublesome notation $\pi\mathcal{P} = \pi$. We call the synthesis (1.3) a *block*. In other words, $\phi_1, \phi_2, \dots, \phi_k$ combined produce the update rule Φ that will generate a Markov chain with invariant measure π . The mapping $x \mapsto \Phi(x)$ evolves paths of this chain starting from anywhere in the state space S to the end of the block.

It is critical that we are able to locate the image $\Phi(S) = \{\Phi(x); x \in S\}$ and construct the block in a way that it can project all paths onto a single point. To this end, most of the ϕ_i are built in a way that yields a monotonicity (e.g. $\phi_i(x) \preceq \phi_i(y)$ when $x \preceq y$) or a coalescence property. The monotonicity property could provide upper and lower paths that make possible the identification of a superset of $\Phi(S)$ after carrying out the updates ϕ_i only for these sandwiching paths while the coalescence property makes possible the event of a block being coalescent, i.e. $\Phi(S)$ is a singleton.

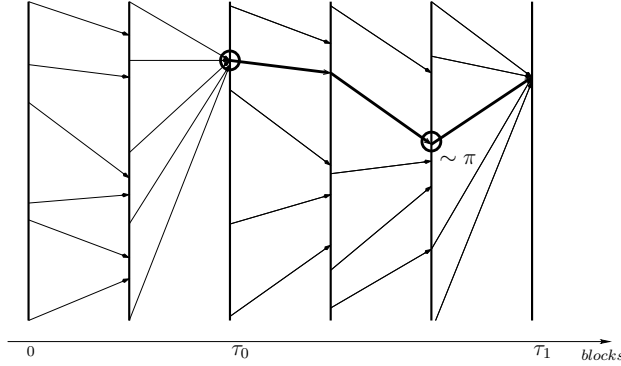


Figure 1.1: A realisation of successive independent Φ -blocks. The thick trajectory represents the coalescence path initiated with the singleton outputted by the first successful block. The instances of this path one block-step before every successful block are iid observations from π .

The Ro-CFTP protocol involves the repeated, forward realisation of a number of independent Φ -blocks and the consideration of their image $\Phi(S)$. In most cases $\Phi(S)$ is difficult to determine and, subsequently, it is impossible to decide whether the coalescence event $\{\Phi(S) \text{ is a singleton}\}$ has taken place or not after running a block. Instead, it can be easier to determine whether a *sub-event*, say D , of the actual coalescence event has occurred or not. For the Ro-CFTP method to be possible it is then necessary that:

$$\mathbb{P}[D] =: \rho > 0. \quad (1.5)$$

Assume that Φ^i is the i -th iteration of a block. Let $\mathbf{w}$ be the observation achieved at the first instance of a *successful* block, i.e. a block when D occurred, say at the τ_0 -th iteration ($\mathbf{w} \equiv \Phi^{\tau_0}(S)$). We call the path initiated by the singleton $\mathbf{w}$ of the first successful block and constructed by its updates thereafter the *coalescence*

path. Define:

$$\tau_i = \min \{ j > \tau_{i-1} : \Phi^j \text{ is successful} \}, \quad i \geq 1.$$

Theorem 1.1 $\Phi^{\tau_{i-1}} \circ \dots \circ \Phi^{\tau_0+1} \circ \Phi^{\tau_0}(S)$, $i \geq 1$, are independent draws from π .

Proof: In the Appendix.

In other words, the instances of the coalescence path one block-step before every successful block are iid observations from π , see Fig.1.1. Thus, for each block we need to know just the location of the coalescence path and the status of the block (successful or not).

An intuitive proof of the Ro-CFTP protocol is given in Wilson (2000b) and an algebraic one in Breyer and Roberts (2001). Our presentation of the Ro-CFTP algorithm is a little more general than in the first appearance of the algorithm (Wilson, 2000b) since we shall need to be able to monitor some arbitrary sub-event of the actual coalescence event for determining whether a block is successful or not. To demonstrate the validity of this compromise we have included in the Appendix a sketch of the proof of Theorem 1.1 in line with the one given in Breyer and Roberts (2001) except for some small changes to apply to our case.

A common situation when we resort to a sub-event of $\{\Phi(S) \text{ is a singleton}\}$ is when we can only locate a set $\Phi' \supseteq \Phi(S)$. In this case $D \equiv \{\Phi' \text{ is a singleton}\}$. Provided that marginally the updates ϕ_i are Markovian (as it will be the case for the algorithms to be presented), if we can determine that a block will be unsuccessful before running all the steps within that block it is then not necessary to continue keeping track of all the paths started from S ; it suffices to update only the coalescence path with respect to the distributions $P_i(x, dy) := \mathbb{P}[\phi_i(x) \in dy]$ for the single current location x of the coalescence path from the time instance when failure was inspected to the end of the block. This follows from the simple

consideration that the coalescence path marginally is updated according to the P_i 's anyway and that for an unsuccessful block the Ro-CFTP algorithm needs only the update of the coalescence path.

In summary, to sample from π we run the blocks and keep track of the coalescence path. The instances of this path one block before each successful one constitute iid draws from π . The algorithm runs in finite time (in terms of the iterations of the blocks needed) since from (1.5) it is clear that the number of blocks necessary to get n draws from π is distributed according to a negative binomial distribution with mean $(n+1)/\rho$. The term One-Shot CFTP refers to the specific structure we impose on the block; we attempt coalescence only once at the last step of the block. In the case of TN s, an appropriate Independence sampler step at the beginning of the block forces paths starting from anywhere in the state space into a bounded set while a number of Gibbs sampler steps bring the paths into a small, identifiable hyper-rectangle ensuring that the final coalescence step has big probability of being successful. The construction of a block of the One-Shot algorithm on TN s will be presented in detail in the sections that follow. Fig.1.3 at pg.81 shows the phases in the formation of such a block. The meaning of the symbols on the graph is explained in the sequel.

1.3 The Monotone Independence Coupler

The Monotone Independence Coupler (MIC) exploits the MCMC technique of Independence sampling to turn stochastic domination into almost surely domination and raise at the same time a coalescence event of positive probability. It is presented in Theorem 1.2. Its relevance to the Gibbs sampler and TN s is explained in Section 1.3.1. The *Independence Coupler* has been proposed as a coupling method in Breyer and Roberts (2000); the additional adjective emphasises the monotonicity property induced by a certain version of this coupling technique that we devised.

We begin by presenting the Independence Coupler. Let $P(x, dy)$ be a probability kernel on $X \times \mathcal{E}$ which for any x in some set X defines a probability measure on a sufficiently regular measurable space $(E, \mathcal{E})$. Assume that $p(x, y)$ is the density of $P(x, dy)$ w.r.t. some reference measure on $(E, \mathcal{E})$ and q some other density w.r.t. the same measure. Let $x \mapsto F_x$, $x \in X$, be a random function that couples $\{P(x, \cdot); x \in X\}$, i.e. for each $x \in X$, $P[F_x \in dy] = P(x, dy)$. Draw $Y \sim q$ and $U \sim \text{Un}(0, 1)$ independently. Consider the modification:

$$C \circ F_x = \begin{cases} Y, & \text{if } p(x, F_x) q(Y) U \leq p(x, Y) q(F_x) \\ F_x, & \text{otherwise} \end{cases}, \quad x \in X.$$

Then $x \mapsto C \circ F_x$ is a coupling of $\{P(x, \cdot); x \in X\}$. The proof is given in Breyer and Roberts (2000). The intuition is that $C \circ F_x$ is derived after drawing F_x from $p(x, \cdot)$ and then performing an Independence sampler step with $p(x, \cdot)$ as the stationary density.

MIC is a special case of the Independence Coupler. It exploits a property sufficient to induce stochastic ordering.

Proposition 1.1 *Let $P(x, dy)$ be a probability kernel on $X \times \mathcal{B}(\mathbf{R})$, $X \subseteq \mathbf{R}$, and $p(x, y)$ its density w.r.t. some reference measure. Assume that p has the properties:*

- (i) *all $p(x, \cdot)$, $x \in X$, have common support $S \subseteq \mathbf{R}$.*
- (ii) *$p(x_2, y)/p(x_1, y)$ is increasing in $y \in S$ when $x_1 \leq x_2$.*

Then P is monotone, that is if D_x is the distribution function that corresponds to $P(x, dy)$, $x \in X$, then for $x_1 \leq x_2$, $D_{x_1} \geq D_{x_2}$.

Proof:

Let $x_1 \leq x_2$ be elements of S . Consider the set $A = \{y \in \mathbf{R} : p(x_1, y) > p(x_2, y)\}$. In case that $A = \emptyset$ and there exists t such that $D_{x_1}(t) < D_{x_2}(t)$ then necessarily $D_{x_1}(\infty) < D_{x_2}(\infty)$ which cannot be true. In case that $A \neq \emptyset$ let $\delta = \sup A$. If $\delta = \infty$, then $p(x_1, y) > p(x_2, y)$ for all $y \in S$, so $D_{x_1} \geq D_{x_2}$. Else if $\delta < \infty$ and

there exists $t \geq \delta$ such that $D_{x_1}(t) < D_{x_2}(t)$ then $D_{x_1}(\infty) < D_{x_2}(\infty)$.

□

We can now present MIC. It operates on monotone probability kernels to induce coalescence.

Theorem 1.2 *Consider a probability kernel $P(x, dy)$ on $X \times \mathcal{B}(\mathbf{R})$, $X \subseteq \mathbf{R}$, with density $p(x, y)$ that satisfies properties (i) and (ii) of Proposition 1.1. Assume that F_x is a random map that couples $\{P(x, \cdot); x \in X\}$ such that $F_{x_1} \leq F_{x_2}$ for $x_1 \leq x_2$. Get some $x^* \in X$ and draw $Y \sim P(x^*, dy)$ and $U \sim \text{Un}(0, 1)$ independently. Define the modified random map:*

$$C \circ F_x = \begin{cases} Y, & \text{if } p(x, F_x) p(x^*, Y) U \leq p(x, Y) p(x^*, F_x) \\ F_x, & \text{otherwise} \end{cases}. \quad (1.6)$$

Then:

- (i) $C \circ F_x$ couples $\{P(x, \cdot); x \in X\}$.
- (ii) if $x_1 \leq x_2$ then $C \circ F_{x_1} \leq C \circ F_{x_2}$.

Proof: In the Appendix.

Note that $p(x, \cdot)$ need only be known up to a normalisation constant. It is straightforward that if there exist $x_l, x_u \in X$ such that $x_l \leq x \leq x_u$ for any $x \in X$, then $C \circ F_{x_l} \leq C \circ F_x \leq C \circ F_{x_u}$, so after realising $C \circ F_x$ only for two x 's we can obtain bounds for the updates of all points in X . Clearly, when X is bounded the map $x \mapsto C \circ F_x$ has positive probability of being coalescent. Since $C \circ F_x$ remains increasing in x , MIC can be applied repeatedly using each time independent proposed variates; for example $C \circ C \circ F_x \sim P(x, dy)$ will be increasing in x . The effect of MIC depends heavily on using a proposal $p(x^*, \cdot)$ of the same type with $p(x, \cdot)$. In different case, result (ii) of Theorem 1.2 would not be generally true.

A firm prerequisite for Theorem 1.2 is that it is possible to obtain a-priori a random map F_x that realises the stochastic ordering of the kernel $P(x, dy)$. Given that, MIC is easy to apply. Although in theory this is always possible using the inverse cdf method (see for example ch.1 of Thorisson, 2000) in practise difficulties related with the availability of certain inverse cdfs are often encountered.

There are also other coupling methods that induce coalescence. The splitting technique (see for instance ch.1 of Thorisson, 2000) obtains the largest possible probability for the coalescence event and so does the layered multi-shift coupler (Wilson, 2000a). However, both methods have limitations in their use.

The layered multi-shift coupler requires that for any $x \in X$, and $F_x \sim P(x, dy)$, the distribution of $F_x - x$ to be independent of x , a strong random-walk assumption. In the case of TNs , the layer multi-shift coupler cannot be applied.

For the splitting technique, one needs to find $r(y) := \inf_{x \in C} p(x, y)$, $y \in \mathbf{R}$, and its integral $\int_{\mathbf{R}} r(y) dy$ for a suitably chosen small set $C \subseteq X$ and simulate from the densities proportional to $r(\cdot)$ and $p(x, \cdot) - r(\cdot)$; the calculations and simulations required create difficulties at the application of the method. MIC is computationally more efficient than the splitting construction (at least in the case of TNs) and benefits from an automatic implementation which does not require application-specific tuning. In fact it can be thought of as a kind of random splitting construction where the smoothness of the transition kernel $P(x, \cdot)$ as a function of x automatically constructs (in a random fashion) an appropriate small set for coupling purposes.

Example: Consider the gaussian family $\{\mathcal{N}(\mu, \sigma^2); \mu \in M \subset \mathbf{R}\}$ for some constant variance σ^2 . It can be equivalently represented by the (unnormalised) kernel $p(\mu, y) = \exp\{-(y - \mu)^2/2\sigma^2\}$, $y \in \mathbf{R}$, which satisfies both conditions of Proposition 1.1. A trivial way to realise the stochastic ordering of the specified family w.r.t. μ is by setting $F_\mu = \mu + \sigma Z$ for a common $Z \sim \mathcal{N}(0, 1)$. Theorem 1.2 now ascertains that if some μ^* is picked up and a variate $Y \sim \mathcal{N}(\mu^*, \sigma^2)$ is

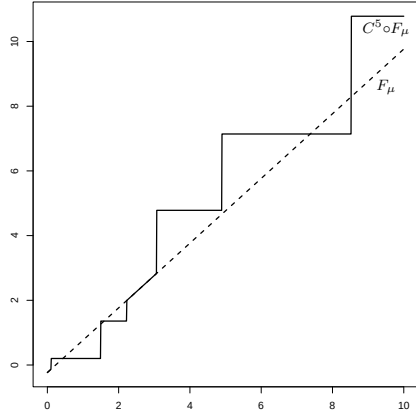


Figure 1.2: The dashed line depicts the map $F_\mu = \mu + Z \sim \mathcal{N}(\mu, 1)$, $\mu \in [0, 10]$, whereas the solid line the modified map $C^5 \circ F_\mu$ after a 5-fold implementation of MIC. $C^5 \circ F_\mu \sim \mathcal{N}(\mu, 1)$ remains increasing in μ .

proposed the modified map $C \circ F_\mu$ will remain increasing in μ . The same procedure can be applied repeatedly with independent Y 's (Fig.1.2).

1.3.1 MIC in the context of TN s

In the Perfect Simulation algorithms to be presented MIC is applied to the univariate updates of a Gibbs sampler step at the end of the block of the One-Shot algorithm. According to the One-Shot rule, this final step is the only one within the block that attempts to merge the paths started from everywhere in the state space at the beginning of the block. The analysis that follows is thus focused on coalescence.

To update the i -th element of some $\mathbf{x}$ according to the Gibbs scheme on $TN_d(\mathbf{0}, Q^{-1})|_{\mathbf{a}}^b$ we need to sample from the univariate law $TN_1(\mu_i(\mathbf{x}_{-i}), 1)|_{\mathbf{a}_i}^{b_i}$ for $\mu_i(\mathbf{x}_{-i}) = -\sum_{j:j \neq i} q_{ij}x_j$. Certainly, we will have to update a set of points so we need to devise a random map $\mu \mapsto F(\mu; \alpha, \beta)$ that couples $TN_1(\mu, 1)|_{\alpha}^{\beta}$ for μ in

some appropriate interval $M = [\mu_l, \mu_u]$. We will give the main idea for the generation of $F(\mu; \alpha, \beta)$ and the way that we check if $\mu \mapsto F(\mu; \alpha, \beta)$ is coalescent for the case of both algorithms (the case-specific and the pure algorithm). Details on the incorporation of this univariate update rule in the general frame of the algorithms are given in the following sections. Note that the family $\{TN_1(\mu, 1)|_{\alpha}^{\beta}; \mu \in M\}$ can be represented by the kernel:

$$p(\mu, y) = \exp \left\{ -\frac{(y - \mu)^2}{2} \right\} \mathbb{I}_{[\alpha, \beta]}(y), \quad y \in \mathbf{R}, \quad \text{for } \mu \in M, \quad (1.7)$$

which satisfies both conditions of Proposition 1.1.

In the case-specific algorithm the use of the cdf of the $\mathcal{N}(0, 1)$ and its inverse can yield a random map $F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ increasing in $\mu \in M$. Theorem 1.2 now ascertains that we can produce a modification $C \circ F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ that remains increasing in μ and has the additional property of inducing coalescence. Clearly, if $C \circ F(\mu_i; \alpha, \beta) = C \circ F(\mu_u; \alpha, \beta)$ then coalescence has happened.

In the pure algorithm, we will produce a random map $F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ piece-wise increasing in μ . Considering each piece separately we can use MIC and obtain a modification $C \circ F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ that remains piece-wise increasing in μ . Recall that the modified map is generated after proposing a $Y \sim p(\mu^*, \cdot)$ for some $\mu^* \in \mathbf{R}$ instead of $F(\mu; \alpha, \beta)$. It is important that we use the same Y for all the increasing pieces since this makes it possible for $C \circ F(\mu; \alpha, \beta)$ to be coalescent. Clearly, we need to determine $C \circ F(\mu; \alpha, \beta)$ explicitly only for the μ 's at the thresholds of the increasing pieces of $F(\mu; \alpha, \beta)$ to decide if coalescence occurred.

1.4 The case-specific algorithm

This is the case when the cdf Φ of the $\mathcal{N}(0, 1)$ and its inverse Φ^{-1} are used. Thus, it involves an algorithm that does not readily extend beyond the gaussian family.

The One-Shot algorithm block involves three phases. For each phase we are going to give the update rule and describe the way we locate in practise the position of the sample paths having started from everywhere in the support of the target density π given at (1.1) after carrying out the certain update. Notice that we always have to monitor the updates of the coalescence path and therefore apply explicitly the provided update rules for the current location of this single trajectory.

In what follows in this section and Section 1.5 it is necessary that we are provided with a fast generator of draws from the univariate distributions $TN_1(\mu, 1)|_\alpha^\beta$ and the one with density:

$$q(x) \propto \exp\left(-\frac{|x|}{\epsilon}\right), \quad x \in [\alpha, \beta], \quad (1.8)$$

for $\epsilon > 0$ and $\alpha, \beta \in \bar{\mathbf{R}}$, $\alpha < \beta$. We denote this distribution by $\text{Exp}(\epsilon)|_\alpha^\beta$. An efficient and numerically stable algorithm for sampling from the $TN_1(\mu, 1)|_\alpha^\beta$ for any values of the parameters is presented in the Appendix. For the case of the $\text{Exp}(\epsilon)|_\alpha^\beta$ a simple algorithm involves well-known methods for simulating truncated univariate exponential densities, so we avoid the details. From now on it is assumed that we can call on the subroutines $\text{tr_normal}(\alpha, \beta, \mu)$ and $\text{tr_exp}(\alpha, \beta, \epsilon)$ to draw from these two univariate distributions.

Phase 1: Forcing paths into a compact set

Consider the density:

$$q(\mathbf{x}) \propto \exp\left(-\sum_{i=1}^d \frac{|x_i|}{\epsilon}\right) \cdot \mathbb{I}_S(\mathbf{x}), \quad \mathbf{x} \in \mathbf{R}^d,$$

where $S = \{\mathbf{x} \in \mathbf{R}^d : \mathbf{a} \preceq \mathbf{x} \preceq \mathbf{b}\}$ is identical to the support of π and $\epsilon > 0$. Sampling from q is straightforward after calling d times on the $\text{tr_exp}(a_i, b_i, \epsilon)$ method. Assume that $B \sim q$ and $U \sim \text{Un}(0, 1)$ independently. Consider the

random update function:

$$\phi_{\mathcal{I}}(\mathbf{x}) = \begin{cases} B, & \text{if } \pi(\mathbf{x})q(B)U \leq \pi(B)q(\mathbf{x}) \\ \mathbf{x}, & \text{otherwise} \end{cases}, \quad \mathbf{x} \in S. \quad (1.9)$$

This is just an Independence sampler step known to preserve π . It is easy to prove that if the support of π allows the consideration of the limit $|\mathbf{x}| \rightarrow \infty$ then:

$$\lim_{|\mathbf{x}| \rightarrow \infty} \frac{\pi(\mathbf{x})}{q(\mathbf{x})} = 0, \quad (1.10)$$

i.e. q is heavier tailed than π . From (1.9), (1.10), it is now clear that points sufficiently distanced from the bulk of the target distribution π accept B rendering the image $\phi_{\mathcal{I}}(S)$ a bounded set. More analytically, assume that R_B is the set of points not accepting B . Then:

$$R_B = \left\{ \mathbf{x} \in S : U > \frac{\pi(B)q(\mathbf{x})}{\pi(\mathbf{x})q(B)} \right\} = \left\{ \mathbf{x} \in S : \epsilon \mathbf{x}^\top Q \mathbf{x} - 2 \sum_{i=1}^d |x_i| < c \right\},$$

for $c = 2\epsilon \log U - 2 \sum |B_i| + \epsilon B^\top Q B$. The following proposition helps in locating a superset of R_B .

Proposition 1.2 *Let $V \in \mathbf{R}^{d \times d}$ be a symmetric, positive definite matrix. Then, for any $\mathbf{x} \in \mathbf{R}^d$, $\mathbf{x}^\top V \mathbf{x} \geq \lambda_V \mathbf{x}^\top \mathbf{x}$, where $\lambda_V > 0$ is the minimum eigenvalue of V .*

Proof: In the Appendix.

This Proposition yields that $R_B \subseteq \{\mathbf{x} \in S : \lambda_Q \epsilon \mathbf{x}^\top \mathbf{x} - 2 \sum |x_i| < c\}$, where λ_Q is the minimum eigenvalue of Q , and after setting $\epsilon = 1/\lambda_Q$ and adjusting c for this value of ϵ :

$$\begin{aligned} R_B \subseteq \{\mathbf{x} \in S : \mathbf{x}^\top \mathbf{x} - 2 \sum |x_i| < c\} &= \{\mathbf{x} \in S : \sum (|x_i| - 1)^2 < c + d\} \subseteq \\ &\subseteq \{\mathbf{x} \in S : |x_i| \leq \sqrt{c + d} + 1, 1 \leq i \leq d\}. \end{aligned}$$

If $c + d < 0$ then $R_B = \emptyset$ and all the state space is projected onto B . Consider the sets $R^i = \{x \in [a_i, b_i] : |x| \leq \sqrt{c+d} + 1\}$, $1 \leq i \leq d$. Clearly, the image $\phi_{\mathcal{I}}(S)$ is a subset of the hyper-rectangle, call it H , determined by the extreme points $\mathbf{x}_l$ and $\mathbf{x}_u$ with individual elements $x_{l,i} = \min\{\psi_{l,i}, B_i\}$ and $x_{u,i} = \max\{\psi_{u,i}, B_i\}$ for $\psi_{l,i} = \inf R^i$ and $\psi_{u,i} = \sup R^i$, $1 \leq i \leq d$. As shown at Phase 1 of Fig.1.3 paths

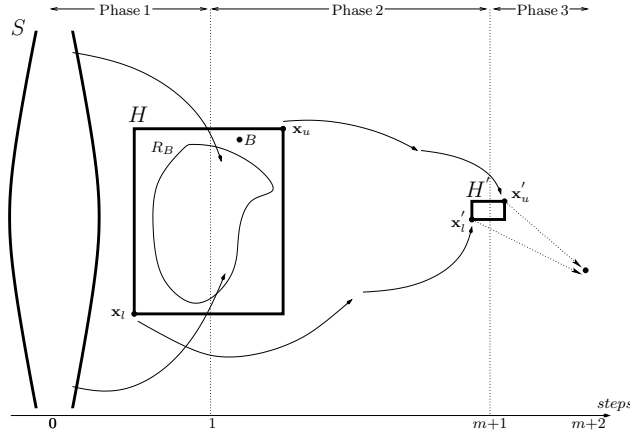


Figure 1.3: A Block of the One-Shot Algorithm on TNs. An Independence sampler step (Phase 1) brings paths into the bounded set H , m Gibbs steps (Phase 2) force paths into the smaller set H' and the attempt for coalescence takes place at an additional Gibbs step (Phase 3).

from everywhere in S are located into the identifiable bounded set H after a single Independence sampler step.

Phase 2: Bridging the gap

This part of the block makes use of a number of Gibbs sampler steps. The objective is to force paths from H into some smaller hyper-rectangle H' . The critical point is that when the precision matrix Q is Stieltjes the Gibbs sampler on π exhibits a monotonicity property that makes necessary (for finding H') the monitoring of only

two paths, an upper and a lower starting from $\mathbf{x}_u$ and $\mathbf{x}_l$ respectively. Recall that if $\mathbf{X} \sim \pi$ then $X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i} \sim TN_1(\mu_i(\mathbf{x}_{-i}), 1)|_{a_i}^{b_i}$ for $\mu_i(\mathbf{x}_{-i}) = -\sum_{j:j \neq i} q_{ij}x_j$. To sample from this conditional density we resort to the inverse cdf method. The inverse cdf of the $TN_1(\mu, 1)|_{a_i}^{b_i}$ is:

$$D^{-1}(u; \mu, \alpha, \beta) := \Phi^{-1}\{u\Phi(\beta - \mu) + (1 - u)\Phi(\alpha - \mu)\} + \mu, \quad u \in [0, 1], \quad (1.11)$$

and is increasing in μ , i.e. $D^{-1}(\cdot; \mu_1, \alpha, \beta) \leq D^{-1}(\cdot; \mu_2, \alpha, \beta)$ for $\mu_1 \leq \mu_2$.

The update of the i -th element of $\mathbf{x}$ under the Gibbs scheme can be given by the random map:

$$\phi_g^i(\mathbf{x}) := (x_1, \dots, x_{i-1}, D^{-1}(\xi; \mu_i(\mathbf{x}_{-i}), a_i, b_i), x_{i+1}, \dots, x_d),$$

for $\xi \sim \text{Un}(0, 1)$. When $\mathbf{x} \preceq \mathbf{y}$ it is true that $\mu_i(\mathbf{x}_{-i}) \leq \mu_i(\mathbf{y}_{-i})$, so $\phi_g^i(\mathbf{x}) \preceq \phi_g^i(\mathbf{y})$.

A full sweep of the Gibbs sampler will be:

$$\phi_g(\mathbf{x}) := \phi_g^d \circ \phi_g^{d-1} \circ \dots \circ \phi_g^2 \circ \phi_g^1(\mathbf{x}),$$

and clearly $\phi_g(\mathbf{x}) \preceq \phi_g(\mathbf{y})$ for $\mathbf{x} \preceq \mathbf{y}$. In general, many iterations, say m , of the Gibbs sampler will be carried out. The overall random map for this part of the block will be:

$$\phi_{\mathcal{G}}(\mathbf{x}) := \phi_g \circ \phi_g \circ \dots \circ \phi_g(\mathbf{x}) = \phi_g^{(m)}(\mathbf{x}), \quad \mathbf{x} \in H. \quad (1.12)$$

Since $\phi_{\mathcal{G}}(\mathbf{x}_l) \preceq \phi_{\mathcal{G}}(\mathbf{x}) \preceq \phi_{\mathcal{G}}(\mathbf{x}_u)$ for all $\mathbf{x} \in H$, when the Gibbs steps are finished all paths have been forced into the tractable hyper-rectangle H' determined by $\mathbf{x}'_l = \phi_{\mathcal{G}}(\mathbf{x}_l)$ and $\mathbf{x}'_u = \phi_{\mathcal{G}}(\mathbf{x}_u)$. That means that $\phi_{\mathcal{G}} \circ \phi_{\mathcal{I}}(S) \subseteq H' \equiv \phi_{\mathcal{G}}(H)$, see Fig.1.3 - Phase 2.

Phase 3 : One-Shot coalescence

We proceed with one more Gibbs step adjoined with an effort to achieve coalescence of the paths. As pointed out in Section 1.3.1, the idea is to use the Monotone Independence Coupler (MIC) at each of the d univariate updates of a full Gibbs sweep. Coalescence is declared when all d attempts are successful. An alternative choice would be the splitting technique, its implementation however would have been much more demanding.

To perform a univariate update of the Gibbs sampler for all points in some certain set we need to draw from $TN_1(\mu, 1)|_\alpha^2$, or equivalently from the kernel $p(\mu, \cdot)$ given in (1.7), for any μ in some set $M = [\mu_l, \mu_u]$ determined by the extreme points of the set and appropriate α, β . For instance, at the first univariate update we need to construct a map for all $\mathbf{x} \in H'$, so $\mu_l = \mu_1(\mathbf{x}'_{l,-1})$ and $\mu_u = \mu_1(\mathbf{x}'_{u,-1})$.

MIC involves two phases. First we construct a map $F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ exactly as in the previous phase of the algorithm; we set $F(\mu; \alpha, \beta) = D^{-1}(\xi; \mu, \alpha, \beta)$ for $\xi \sim \text{Un}(0, 1)$ and D^{-1} defined in (1.11). Recall that $F(\mu; \alpha, \beta)$ is increasing in μ . Then we pick up some $\mu^* \in \mathbf{R}$ and draw $Y \sim p(\mu^*, \cdot)$ and $U \sim \text{Un}(0, 1)$ independently. Consider the modified map (for simplicity we write F_μ instead of $F(\mu; \alpha, \beta)$):

$$C \circ F_\mu = \begin{cases} Y, & \text{if } p(\mu, F_\mu) p(\mu^*, Y) U \leq p(\mu, Y) p(\mu^*, F_\mu) \\ F_\mu, & \text{otherwise} \end{cases}.$$

We have already pointed out that the kernel p satisfies both conditions of Proposition 1.1, so Theorem 1.2 guarantees that $C \circ F_\mu \sim p(\mu, \cdot)$ and is increasing in μ . At the actual implementation of the algorithm we choose $\mu^* = (\mu_l + \mu_u)/2$ so that the probability of coalescence is relatively high and evaluate $C \circ F_\mu$ only for $\mu = \mu_l$ and $\mu = \mu_u$. These two values are the updates of a certain co-ordinate of

the sandwiching paths and will be used at the next univariate update to find the new boundaries μ_l, μ_u .

For the general algorithm, we perform the i -th univariate update of the final Gibbs step via the random map:

$$\phi_c^i(\mathbf{x}) := (x_1, \dots, x_{i-1}, C \circ F(\mu_i(\mathbf{x}_{-i}); a_i, b_i), x_{i+1}, \dots, x_d).$$

For $\mathbf{x} \preceq \mathbf{y}$ it is clear that $\mu_i(\mathbf{x}_{-i}) \leq \mu_i(\mathbf{y}_{-i})$, so $\phi_c^i(\mathbf{x}) \preceq \phi_c^i(\mathbf{y})$. The complete update rule is:

$$\phi_c(\mathbf{x}) := \phi_c^d \circ \phi_c^{d-1} \circ \dots \circ \phi_c^2 \circ \phi_c^1(\mathbf{x}), \quad \mathbf{x} \in H', \quad (1.13)$$

and is increasing in $\mathbf{x}$. Someone needs to consider its effect only on the two extreme paths to determine if coalescence is achieved. If at some intermediate univariate update the corresponding elements of the extreme paths do not couple we declare the block unsuccessful and update only the coalescence path according to the transition kernel determined by the Gibbs sampler (see the discussion in Section 1.2).

Synthesis

The compound update rule for a block will be $\Phi = \phi_c \circ \phi_g \circ \phi_I$. It preserves π since all its individual update functions defined in (1.9), (1.12) and (1.13) do. Because of the Independence sampler step and the fact that we can only locate a superset H of $\phi_I(S)$ it is obvious that $\Phi(S)$ is intractable. As already shown, we can easily obtain some $\Phi' \supseteq \Phi(S)$ determined as $\Phi' = \phi_c(\phi_g(H))$. Φ' can be a singleton with positive probability increasing in the number of Gibbs steps within each block.

It is important that the finite accuracy of the computing systems restricts the availability of the cdf Φ of the standard normal distribution and its inverse Φ^{-1} .

The double-precision GSL library for the C-language returns values for $\Phi(x)$ only when $|x| < 38$. For instance, when the inverse cdf formula (1.11) is used for the $TN_1(0, 1)|_{40}^{50}$ it returns $+\infty$. If it happens that a value $\Phi(x)$ for x outside of the range $[-38, 38]$ is needed we instantly declare the block unsuccessful (we cannot realise the monotonicity property anymore) and update only the coalescence path until the end of the block. This action is validated by considering the event $D = \{\Phi' \text{ is a singleton}\} \cap \{\text{the block does not require a value } \Phi(x) \text{ for } |x| > 38\}$ and use it as the sub-event of the actual coalescence event $\{\Phi(S) \text{ is a singleton}\}$ at the demonstration of the Ro-CFTP protocol in Section 1.2.

1.5 The pure algorithm

Recall that the target distribution is π given in (1.1) for $\mu = \mathbf{0}$, support S and Stieltjes precision matrix $Q = (q_{ij})$ with $q_{ii} = 1$, $1 \leq i \leq d$. This time the use of Φ or Φ^{-1} is not considered yielding the prospect of implementing the resulted algorithm in a context more general than that of TN s. The phases in the construction of a block are the same as for the case-specific algorithm. Again, for each phase we give the official update rule and describe the way we identify in practise the position of the paths after that update has taken place. Note that the information we need at the end of a block is the location of the coalescence path (so it is always presumed that we explicitly apply the update rules for it) and whether it is successful or not.

Phase 1: Forcing the paths into a compact set

It is identical to Phase 1 of the case-specific algorithm. It yields the random map ϕ_T (1.9) that projects all elements of S onto a hyper-rectangle H determined

by two identifiable points $\mathbf{x}_l$ and $\mathbf{x}_u$, see Fig.1.3 - Phase 1.

Phase 2: Bridging the gap

Assume that a path is located at some $\mathbf{x} \in S$ and the update of the i -th element of $\mathbf{x}$ is about to take place via Gibbs sampler. The full conditional distribution of the i -th element of a random vector distributed according to π is given in (1.2). The case-specific algorithm exploited the fact that the inverse cdf of this univariate distribution is increasing in $\mathbf{x}$ and implemented the Gibbs steps only for the sandwiching paths found at $\mathbf{x}_l, \mathbf{x}_u$ before the effect of the Gibbs sampler. Without Φ, Φ^{-1} realising the stochastic ordering of the family $\{TN_1(\mu, 1)|_\alpha^\beta; \mu \in M \subseteq \mathbf{R}\}$ is impossible. Instead we opt for an adaptive rejection sampling scheme. It yields two bounding paths that set off from $\mathbf{x}_l$ and $\mathbf{x}_u$, squeeze all paths between them but are technically constructed, i.e. they do not represent actual realisations of the underlying Markov chain but a selection of instances among several paths.

Let ϕ_g^i denote the update rule under the i -th univariate Gibbs step. To construct ϕ_g^i two methods are considered. The first (Fig.1.4) provides a random map $R_1(\mu; \alpha, \beta) \sim TN_1(\mu, 1)|_\alpha^\beta$ for all $\mu \in [\mu_l, \mu_u] \subseteq [\alpha, \beta]$. It involves rejection sampling that uses as envelopes the corresponding untruncated densities. If $(Z + \mu) \sim \mathcal{N}(\mu, 1)$ lies in $[\alpha, \beta]$ then $Z + \mu$ is accepted as a draw for $TN_1(\mu, 1)|_\alpha^\beta$. It uses the same Z for all μ 's and is adaptive since the μ 's that accept $Z + \mu$ are located and a subset of $[\mu_l, \mu_u]$ is considered for the rest of the algorithm. The algorithm terminates when $R_1(\mu; \alpha, \beta)$ is constructed for all $\mu \in [\mu_l, \mu_u]$.

The second method generates a map $R_2(\mu; \alpha, \beta) \sim TN_1(\mu, 1)|_\alpha^\beta$ for any μ in some interval $[\mu_l, \mu_u]$ when $\mu_u \leq \alpha$ or $\mu_l \geq \beta$. It is a case of *Rejection Coupler* (Murdoch and Green, 1998a). The unnormalised density of the $TN_1(\mu, 1)|_\alpha^\beta$

```

set  $\mu_{\min} = \mu_l, \mu_{\max} = \mu_u$ 
while ( $\mu_{\min} \neq \mu_{\max}$ ) repeat
  sample increment  $Z \sim \mathcal{N}(0, 1)$ 
  if ( $Z + \mu_{\min} \leq \beta$  and  $Z + \mu_{\max} \geq \alpha$ )
    if ( $Z + \mu_{\min} \geq \alpha$  and  $Z + \mu_{\max} \leq \beta$ )
      store  $(\mu_{\min}, Z + \mu_{\min}), (\mu_{\max}, Z + \mu_{\max})$ , set  $\mu_{\max} = \mu_{\min}$ 
    else if ( $Z + \mu_{\min} \geq \alpha$ )
      store  $(\mu_{\min}, Z + \mu_{\min}), (\beta - Z, \beta)$ , set  $\mu_{\min} = \beta - Z$ 
    else if ( $Z + \mu_{\max} \leq \beta$ )
      store  $(\alpha - Z, \alpha), (\mu_{\max}, \mu_{\max} + Z)$ , set  $\mu_{\max} = \alpha - Z$ 
return stored points

```

Figure 1.4: The algorithm that generates a piece-wise increasing random map $\mu \mapsto R_1(\mu; \alpha, \beta)$ that couples $\{TN_1(\mu, 1)|_{\alpha}^{\beta}; \mu \in [\mu_l, \mu_u] \subseteq [\alpha, \beta]\}$.

```

set  $\mu_{\min} = \mu_l, \mu_{\max} = \mu_u$ 
while ( $\mu_{\min} \neq \mu_{\max}$ ) repeat
  sample  $Y \sim p(\mu_{\max}, \cdot), U \sim \text{Un}(0, 1)$ 
  set  $\mu_{\max} = \min\{\mu : \mu_{\min} \leq \mu \leq \mu_{\max} \text{ and } U \leq p(\mu, Y)/p(\mu_{\max}, Y)\}$ 
  store  $(\mu_{\max}, Y)$ 
return stored points

```

Figure 1.5: The algorithm that generates the piece-wise constant random map $\mu \mapsto R_2(\mu; \alpha, \beta)$ that couples $\{TN_1(\mu, 1)|_{\alpha}^{\beta}; \mu \in [\mu_l, \mu_u] \subseteq (-\infty, \alpha]\}$.

distribution $p(\mu, y) = \exp\{-(y - \mu)^2/2\} \mathbb{I}_{[\alpha, \beta]}(y)$ has the property:

$$\text{if } \mu_1 \leq \mu_2 \leq \alpha \text{ or } \beta \leq \mu_2 \leq \mu_1 \text{ then } p(\mu_1, \cdot) \leq p(\mu_2, \cdot). \quad (1.14)$$

The algorithm that constructs R_2 (for the case when $\mu_u \leq \alpha$) involves adaptive rejection sampling; use $p(\mu_u, \cdot)$ as an envelope for the others $p(\mu, \cdot)$, $\mu \in [\mu_l, \mu_u]$, locate the minimum μ that accepts a proposal from $p(\mu_u, \cdot)$ and set that as the new μ_u . It terminates when a draw from $p(\mu_l, \cdot)$ is obtained. For the case when $\mu_l \geq \beta$ we follow the same procedure with the apparent symmetrical considerations. The pseudo-algorithm in Fig.1.5 is for $\mu_u \leq \alpha$. Fig.1.6 shows a joint realisation of the random functions $R_1(\mu; \alpha, \beta)$, $R_2(\mu; \alpha, \beta)$, considered to be independent between them, for the case when $\alpha = -1$, $\beta = 1$ and $\mu \in [-2, 2]$.

The overall update rule for the i -th univariate update of a Gibbs step is defined

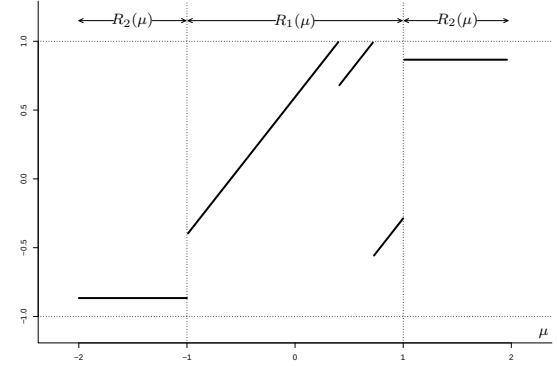


Figure 1.6: A realisation of the random update maps $R_1(\mu; -1, 1) \sim TN_1(\mu, 1)|_{-1}^1$ for $\mu \in [-1, 1]$ and $R_2(\mu; -1, 1) \sim TN_1(\mu, 1)|_{-1}^1$ for $\mu \in [-2, -1] \cup [1, 2]$.

as:

$$\phi_g^i(\mathbf{x}) = \begin{cases} (x_1, \dots, x_{i-1}, R_1(\mu_i(\mathbf{x}_{-i}); a_i, b_i), x_{i+1}, \dots, x_d), & a_i \leq \mu_i(\mathbf{x}_{-i}) \leq b_i \\ (x_1, \dots, x_{i-1}, R_2(\mu_i(\mathbf{x}_{-i}); a_i, b_i), x_{i+1}, \dots, x_d), & \mu_i(\mathbf{x}_{-i}) > b_i \text{ or } < a_i \end{cases},$$

where we have assumed that ϕ_g^i applies to a set points within a bounding rectangle so that it is possible to identify the appropriate bounds μ_l, μ_u for the construction of R_1 and R_2 . For instance at the first univariate update of the first Gibbs step this rectangle is H . As the following update rules take place this bounding area is determined by a pair of bounding paths whose construction is detailed in the next paragraph.

A full Gibbs step is given as:

$$\phi_g(\mathbf{x}) := \phi_g^d \circ \phi_g^{d-1} \circ \dots \circ \phi_g^1(\mathbf{x}).$$

Carrying out m Gibbs steps according to ϕ_g corresponds to using the random map:

$$\phi_{\mathcal{G}}(\mathbf{x}) := \phi_g \circ \phi_g \circ \dots \circ \phi_g(\mathbf{x}) = \phi_g^{(m)}(\mathbf{x}), \quad \mathbf{x} \in H. \quad (1.15)$$

Clearly, $\phi_G(\mathbf{x})$ is not monotone in $\mathbf{x}$ (see Fig.1.6) so locating the position of the paths after applying it is not straightforward. To overcome this inconvenience we develop a pair of bounding paths.

We denote by R the complete update map produced after combining R_1 and R_2 . Let $\mathbf{y}_u^{(i,j)}, \mathbf{y}_l^{(i,j)}$ be the position of the upper and lower bounding paths after the j -th univariate update of the i -th Gibbs step has taken place, $1 \leq i \leq m$, $0 \leq j \leq d$. These paths begin from the edges of the hyper-rectangle H determined by the Independence sampler step, i.e. $\mathbf{y}_u^{(1,0)} \equiv \mathbf{x}_u$, $\mathbf{y}_l^{(1,0)} \equiv \mathbf{x}_l$, and are developed in a way that $\mathbf{y}_u^{(i,j)} (\mathbf{y}_l^{(i,j)})$ is identical to $\mathbf{y}_u^{(i,j-1)} (\mathbf{y}_l^{(i,j-1)})$ except for its j -th element which is given as the $\sup A_{i,j}$ ($\inf A_{i,j}$) for the set $A_{i,j}$ defined as $A_{i,j} = \{R(\mu_j(\mathbf{y}_{-j}); a_j, b_j); \mathbf{y}_l^{(i,j-1)} \preceq \mathbf{y} \preceq \mathbf{y}_u^{(i,j-1)}\}$, under the convention that $(i, d) \equiv (i+1, 0)$.

Essentially, at each univariate update of the Gibbs sampler we locate the minimum and the maximum value of the univariate map R that updates the certain co-ordinate for all the points between the current location of the bounding paths and use these values to update the bounding paths. The definition of R renders the identification of these values an easy task. This procedure yields a tractable hyper-rectangle H' determined by the extreme points $\mathbf{x}'_l := \mathbf{y}_l^{(m,d)}$ and $\mathbf{x}'_u := \mathbf{y}_u^{(m,d)}$ that includes the position of all paths started from H after the effect of these m Gibbs steps.

Phase 3: One-Shot coalescence

As with the case-specific algorithm we try to couple the univariate updates of an additional Gibbs step. In the case of Section 1.4 we were able to construct a random map $F(\mu; \alpha, \beta) \sim TN_1(\mu, 1)|_\alpha^\beta$ (equivalently $F(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ for the kernel p defined in (1.7)) increasing in μ rendering the effect of MIC straightforward. Now we follow a different procedure.

As in the previous phase of this algorithm we combine, if necessary, R_1, R_2 to get the random map $R(\mu; \alpha, \beta) \sim p(\mu, \cdot)$. The generation of R_1, R_2 ensures that R is piece-wise increasing in μ , see Fig.1.6. We choose $\mu^* \in \mathbf{R}$ and draw $Y \sim p(\mu^*, \cdot)$ and $U \sim \text{Un}(0, 1)$ independently. Consider the modified map:

$$C \circ R(\mu; \alpha, \beta) = \begin{cases} Y, & \text{if } \mu \in [\alpha, \beta], \ p(\mu, R_\mu) p(\mu^*, Y) U \leq p(\mu, Y) p(\mu^*, R_\mu) \\ R_\mu, & \text{if } \mu \in [\alpha, \beta], \ p(\mu, R_\mu) p(\mu^*, Y) U > p(\mu, Y) p(\mu^*, R_\mu) \\ R_\mu, & \text{otherwise} \end{cases}, \quad (1.16)$$

where we wrote R_μ instead of $R(\mu; \alpha, \beta)$. The new update rule implements MIC on R only for $\mu \in [\alpha, \beta]$. For $\mu < \alpha$ or $\mu > \beta$ the definition of R itself provides a chance for coalescence without the need for modification. From Theorem 1.2(i) it is clear that $C \circ R(\mu; \alpha, \beta) \sim p(\mu, \cdot)$ for any μ considered. Theorem 1.2(ii) ascertains that $C \circ R(\mu; \alpha, \beta)$ remains piece-wise increasing in μ .

Assume that a pair of bounding paths (the construction of which we explain in the next paragraph) determine that we need to use (1.16) for all $\mu \in [\mu_l, \mu_u]$. A careful consideration of (1.16) indicates that if $\mu_u \leq \alpha$ or $\mu_l \geq \beta$ then we use the Rejection Coupler as the means for coalescence. If $[\mu_l, \mu_u] \subseteq [\alpha, \beta]$ then we apply MIC for $\mu^* = (\mu_l + \mu_u)/2$ and we can determine if the modified update $C \circ R(\mu; \alpha, \beta)$ is coalescent after realising it only for the extreme points of the increasing pieces of $R(\mu; \alpha, \beta)$. In any other case coalescence cannot happen, so we declare the block unsuccessful and update just the coalescence path to the end of the block according to the marginal kernel of the Gibbs sampler.

We do not need to resort to a sophisticated construction of bounding paths in the way described in the previous part of the algorithm. Recall that this is an One-Shot algorithm, so if a univariate Gibbs update fails to merge the element it updates for all points in an area of interest then we declare the block unsuccessful and update only the coalescence path. Clearly, given that after the i -th Gibbs update all univariate updates have been coalescent the area we mention above is

the hyper-rectangle determined by a pair of points equal to $\mathbf{x}'_l, \mathbf{x}'_u$ except for their first i -elements which have been updated and are identical.

In the general context of the algorithm, the i -th update of a Gibbs step is given by the random map:

$$\phi_c^i(\mathbf{x}) := (x_1, \dots, x_{i-1}, C \circ R(\mu_i(\mathbf{x}_{-i}); a_i, b_i), x_{i+1}, \dots, x_d) . \quad (1.17)$$

The complete update rule is:

$$\phi_C(\mathbf{x}) := \phi_c^d \circ \phi_c^{d-1} \circ \dots \circ \phi_c^2 \circ \phi_c^1(\mathbf{x}), \quad \mathbf{x} \in H' . \quad (1.18)$$

Synthesis

The random update map for a complete block of the pure algorithm is the synthesis $\Phi = \phi_C \circ \phi_G \circ \phi_I$, where the constituent update maps are defined in (1.9), (1.15) and (1.18). It has to be mentioned that, in theory, the hyper-rectangle H determined after the Independence sampler step could be extremely large for some blocks and this could create some problems related with the time required to carry out the Rejection Coupler scheme of the subsequent Gibbs steps. For example, it could happen that at a univariate update of the Gibbs step it is required that we construct the map $R_2(\mu; 0, 10) \sim TN_1(\mu, 1)|_0^{10}$ for all $\mu \in [10, 2000]$ which would require considerable time. To avoid such problems we could instantly declare a block unsuccessful when the map $R_2(\mu; \alpha, \beta)$ should be determined for $\mu \in [\mu_l, \mu_u]$ when $\max\{(\mu_u - \beta), (\alpha - \mu_l)\}$ is bigger than some threshold of our choice. For the cases considered in Section 1.7 when we run the algorithms for various target distributions, such problems with the Independence sampler step did not occur so we avoid any further reference to this subtle point.

In total, for the pure algorithm a block is declared successful when the sub-

event $D = \{\text{the bounding processes coalesce}\}$ of the actual coalescence event $\{\Phi(S) \text{ is a singleton}\}$ takes place.

1.6 Extensions

The truncated gaussian distribution (1.1) is used throughout this paper mostly for reasons of demonstrating the applicability of the MIC monotone coupling technique introduced in Section 1.3 and the various continuous state space coupling constructions of Sections 1.4 and 1.5 in a realistic setting. In Section 1.6.2 we give a hint about how similar constructions can be applied to other families of distributions. Before that we make a simple observation about the gaussian case.

1.6.1 Extensions within the gaussian family

Although the preceding algorithms appear to be applicable to a certain class of TNs, those with a Stieltjes precision matrix, a superset of this class can be treated in a similar way. Consider the following two structures of a $d \times d$, real, symmetric, positive definite matrix:

$$\begin{pmatrix} 1 & - & + & \cdots & - & - \\ - & 1 & + & \cdots & - & - \\ + & + & 1 & \cdots & + & + \\ - & - & + & \cdots & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ - & - & + & \cdots & - & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & - & - & \cdots & - & - \\ - & 1 & - & \cdots & - & - \\ - & - & 1 & \cdots & - & - \\ - & - & - & \cdots & - & - \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ - & - & - & \cdots & - & 1 \end{pmatrix}$$

If $\mathbf{X} \sim TN_d(\mathbf{0}, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}}$ with Q having the left side structure then the random vector $\mathbf{X}' = (X_1, X_2, -X_3, X_4, \dots, X_d)$ is distributed according to $TN_d(\mathbf{0}, Q_o^{-1})|_{\mathbf{a}'}^{\mathbf{b}'}$ with $\mathbf{a}', \mathbf{b}'$ identical to $\mathbf{a}, \mathbf{b}$ respectively except for $a'_3 = -b_3$ and $b'_3 = -a_3$ and Q_o a matrix with the right side structure, i.e. a Stieltjes matrix (it can be shown that Q_o remains positive definite). This observation ascertains that our methodology

could be implemented on a class of TN s wider than the one so far considered. In the following definition $r_i : \mathbf{R}^{d \times d} \mapsto \mathbf{R}^{d \times d}$ represents a map which when applied to a $d \times d$ matrix changes the sign of the non-diagonal elements of the i -th row and column, $1 \leq i \leq d$, and r_0 the identity mapping that maps a matrix to itself.

Definition 1.2 *The Positive Association Equivalence Class (PAEC) of truncated gaussian distributions includes distributions identified as $TN_d(\mu, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}}$ for some $d \in \{1, 2, \dots\}$, $\mathbf{a}, \mathbf{b} \in \bar{\mathbf{R}}^d$, $\mu \in \mathbf{R}^d$ and $Q \in \mathbf{R}^{d \times d}$ a symmetric and positive definite matrix for which there exists a sequence of operators $(r_{i_j})_{j=1}^k$ such that:*

$$r_{i_k} \circ r_{i_{k-1}} \circ \dots \circ r_{i_1}(Q)$$

is a Stieltjes matrix, $\{i_1, i_2, \dots, i_k\} \subseteq \{0, 1, 2, \dots, d\}$.

This sign transformation is called *Parity Switching* in Roberts and Tweedie (2006). Clearly, the pure and the case-specific algorithms can be applied to any member of this class of TN s. Actually, even when $TN_d(\mu, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}} \notin PAEC$ it is still possible to call on the same algorithms provided that:

$$\sum_{j:j \neq i} |q_{ij}| < 1 \quad \text{for all } 1 \leq i \leq d. \quad (1.19)$$

While nothing changes for the Independence sampler step a modification of the Gibbs steps is necessary. If the i -th updating of a Gibbs step is taking place with the paths found in the hyper-rectangle determined by some extreme points $\mathbf{x}_l, \mathbf{x}_u$ then a map $F_\mu \sim TN_1(\mu, 1)|_{a_i}^{b_i}$ for all μ in the interval:

$$\left[- \sum_{j:j \neq i} q_{ij} \{x_{l,j} \mathbb{I}(q_{ij} \leq 0) + x_{u,j} \mathbb{I}(q_{ij} > 0)\}, - \sum_{j:j \neq i} q_{ij} \{x_{u,j} \mathbb{I}(q_{ij} \leq 0) + x_{l,j} \mathbb{I}(q_{ij} > 0)\} \right]$$

is constructed in the way described in Phase 2 of the case-specific and pure algorithms. The i -th elements of $\mathbf{x}_l, \mathbf{x}_u$ are updated with the minimum and the

maximum value of the map F_μ respectively. Condition (1.19) guarantees the convergence of the bounding paths constructed by the update rule F_μ .

1.6.2 Further directions for extensions

The Perfect Simulation algorithms of Sections 1.4 and 1.5 can apply in a similar way to any multivariate distribution of arbitrary dimension d for which:

P₁) for any $1 \leq i \leq d$, the density of the distribution of $X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i}$ can be written as $p_i(\epsilon_i(\mathbf{x}_{-i}), \cdot)$ where $\epsilon_i : \mathbf{R}^{d-1} \mapsto \mathbf{R}$, is locally bounded, and

P₂) the density $p_i(\epsilon, \cdot)$ in P₁ is stochastically monotone in ϵ , $1 \leq i \leq d$.

The stochastic ordering in P₂ can be increasing in the parameter, i.e. the ratio $p_i(\epsilon_2, y)/p_i(\epsilon_1, y)$ being increasing in y for $\epsilon_1 \leq \epsilon_2$, as in the case of a TN , or decreasing, i.e. $p_i(\epsilon_2, y)/p_i(\epsilon_1, y)$ decreasing in y for $\epsilon_1 \leq \epsilon_2$, as it will be the case for the examples that follow.

These are not the most general conditions possible, but generality comes at the expense of applicability in this problem, so we concentrate on the family of distributions with properties P₁ and P₂ for which we can demonstrate that the algorithms described so far in the paper can be readily generalised. We remark here that P₁ and P₂ are implied by the stronger though possibly easy to check condition that for a density f :

$$\frac{\partial^2}{\partial x_i \partial x_j} \log f(\mathbf{x})$$

is either non-negative for all pairs pair $1 \leq i < j \leq d$ and all $\mathbf{x}$ or non-positive. That this condition implies P₁ and P₂ is easily verified using the FKG inequality (see for example Karlin, 1968). Note that this condition is trivially satisfied for a TN .

Based on the case of a TN , we will provide the guidelines for the development of an One-Shot algorithm on an arbitrary multivariate density, say f , satisfying

P_1 and P_2 . For the Independence sampler step we need to find a proposal density with heavier tails than f ; the corresponding update function will then force all paths into a compact set, see Phase 1 of Section 1.4. This step might even not be necessary, see the examples that follow.

Assume now that before a Gibbs step that will update the i -th co-ordinate of the paths, all paths are located in a compact set, say G . The Gibbs sampler step will require the construction of a random map $\epsilon \mapsto F_\epsilon$ for all ϵ in the set $E = [\inf_{\mathbf{x} \in G} \epsilon_i(\mathbf{x}_{-i}), \sup_{\mathbf{x} \in G} \epsilon_i(\mathbf{x}_{-i})]$, with the marginal property that $F_\epsilon \sim p_i(\epsilon, \cdot)$ such that it is possible to identify the $\inf_{\epsilon \in E} F_\epsilon, \sup_{\epsilon \in E} F_\epsilon$; these extremes will be the updates of the i -th co-ordinate of some lower and upper bounding paths. If the inverse cdf of $p_i(\epsilon, \cdot)$ is available we easily obtain an F_ϵ monotone in ϵ , see Phase 2 of Section 1.4. Otherwise, it could be possible to construct F_ϵ using some of the advanced coupling techniques of Phase 2 of Section 1.5 (adaptive rejection sampling, Rejection Coupler).

The final phase of the block involves the attempt for coalescence after appropriately modifying an additional Gibbs step. We follow exactly the same rules as for the case of a TN . If the random maps for the univariate updates of the Gibbs step are monotone we use MIC (Phase 3, Section 1.4); else if they are constructed using adaptive rejection sampling or Rejection Coupler we modify them using MIC to induce coalescence or do not modify them and hope for the Rejection Coupler to provide coalescence respectively (Phase 3, Section 1.5).

For this construction to apply, the upper and lower bounding paths developed within the block must converge. This will depend on the specific density under consideration and could be checked theoretically, see Møller (1999) for the case of the multivariate gamma conditional distribution defined in the sequel, or even experimentally, i.e. from preliminary runnings of the Gibbs steps.

The family of distributions with the properties P_1 and P_2 includes many probability models used in the literature; we will now give two examples.

Multivariate gamma conditionals distribution

The so-called autogamma distribution has already been subject of the Perfect Simulation literature, see Møller (1999). It includes as a specific case a bayesian hierarchical model on a dataset on pump reliability originally presented in Gelfand and Smith (1990) and used among others by Rosenthal (1995); Murdoch and Green (1998a); Breyer and Roberts (2001). The density of the autogamma distribution writes as:

$$f(\mathbf{x}) \propto \exp \left\{ - \sum_{i=1}^d \epsilon_i x_i - \sum_{i < j} \epsilon_{ij} x_i x_j \right\} \prod_{i=1}^d x_i^{\delta_i - 1}, \quad \mathbf{x} \in \mathbf{R}^d, \mathbf{x} \succeq \mathbf{0}, \quad (1.20)$$

for $\epsilon_i > 0, \delta_i > 0$ and $\epsilon_{ij} = \epsilon_{ji} \geq 0$. If $\mathbf{X} \sim f$ then :

$$X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i} \sim \Gamma(\delta_i, \epsilon_i(\mathbf{x}_{-i})), \quad \epsilon_i(\mathbf{x}_{-i}) := \epsilon_i + \sum_{j:j \neq i} \epsilon_{ij} x_j. \quad (1.21)$$

$\Gamma(\delta, \epsilon)$ denotes the gamma distribution with mean δ/ϵ . Property P_1 is satisfied for $p_i(\epsilon_i(\mathbf{x}_{-i}), \cdot)$ the density of $\Gamma(\delta_i, \epsilon_i(\mathbf{x}_{-i}))$. Also, $p_i(\epsilon, \cdot)$ is stochastically decreasing in ϵ so P_2 holds.

The Independence sampler step is not necessary. For the Gibbs steps we construct the random map $F_\epsilon \sim p_i(\epsilon, \cdot)$ by setting $F_\epsilon = G/\epsilon$ for $G \sim \Gamma(\delta_i, 1)$. After one Gibbs sweep the upper bounding path initiated at $\{+\infty\}^d$ will have finite values and all paths will have been forced into a compact set; just notice that $\epsilon_i(\mathbf{x}_{-i}) \geq \epsilon_i > 0$, thus the possible values of ϵ are lower bounded and F_ϵ will be upper bounded. As shown in Møller (1999) the bounding paths updated via the supremum and infimum of F_ϵ converge.

Were we given a truncated version of (1.20), we could still build a Perfect Simulation algorithm using the adaptive rejection sampling techniques of Section 1.5.

Multivariate exponential conditionals distribution

For details on this family of distributions see for instance ch.8 of Arnold et al. (1999) and Arnold and Strauss (1988). The density is as follows:

$$f(\mathbf{x}) = \exp \left\{ - \sum_{\mathbf{v} \in \xi_d} \lambda_{\mathbf{v}} \left(\prod_{i=1}^d x_i^{v_i} \right) \right\}, \quad \mathbf{x} \in \mathbf{R}^d, \mathbf{x} \succeq \mathbf{0}, \quad (1.22)$$

where $\xi_d := \{\mathbf{v} \in \mathbf{R}^d : v_i = 0 \text{ or } 1 \text{ for all } 1 \leq i \leq d\}$, $\lambda_{\mathbf{v}}$ for $\mathbf{v} \neq \mathbf{0}$ non-negative parameters and λ_0 a normalizing constant. If $\mathbf{X} \sim f$ then:

$$X_i | \mathbf{X}_{-i} = \mathbf{x}_{-i} \sim \text{Exp}(1/\epsilon_i(\mathbf{x}_{-i})), \quad \epsilon_i(\mathbf{x}_{-i}) := \sum_{\mathbf{v} \in \xi_d, v_i=1} \lambda_{\mathbf{v}} \left(\prod_{j:j \neq i} x_j^{v_j} \right),$$

with $\text{Exp}(1/\epsilon)$, $\epsilon > 0$, the exponential distribution with mean $1/\epsilon$. Property P₁ holds for $p_i(\epsilon_i(\mathbf{x}_{-i}), \cdot)$ the density of $\text{Exp}(1/\epsilon_i(\mathbf{x}_{-i}))$; property P₂ also holds since $p_i(\epsilon, \cdot)$ is stochastically decreasing in ϵ .

It is not simple to give a general rule for the Independence sampler step for any values of the parameters $\lambda_{\mathbf{v}}$. In fact, for a sub-family of (1.22) defined in the sequel this step is not necessary. Assuming that the paths can be projected onto a hyper-rectangle with lower point $\mathbf{x}_l$ of positive co-ordinates and upper point $\mathbf{x}_u$ of co-ordinates $x_{u,j} < \infty$, $1 \leq j \leq d$, the Gibbs steps are then straightforward; we construct the random map $F_{\epsilon} \sim p_i(\epsilon, \cdot)$ for any required ϵ in a bounded interval E using the inverse cdf method:

$$F_{\epsilon} = -\log(1 - \xi)/\epsilon, \quad \xi \sim \text{Un}(0, 1). \quad (1.23)$$

F_{ϵ} is decreasing in ϵ , so we can easily find $\inf_{\epsilon \in E} F_{\epsilon}$ and $\sup_{\epsilon \in E} F_{\epsilon}$ which are used for the update of the i -th co-ordinate of the lower and upper bounding paths initiated from $\mathbf{x}_l$ and $\mathbf{x}_u$ respectively. The coalescence step will be a simple application of MIC.

We return to the Independence sampler step. Consider the densities (see ch.8 of Arnold et al., 1999):

$$f(\mathbf{x}) \propto \exp \left\{ - \sum_{i=1}^d x_i - \delta \prod_{i=1}^d x_i \right\}, \quad \mathbf{x} \in \mathbf{R}^d, \mathbf{x} \succeq \mathbf{0}, \quad (1.24)$$

generated by (1.22) under the specifications $\lambda_{\mathbf{v}} = \delta > 0$ for $\mathbf{v} = (1, 1, \dots, 1)$, $\lambda_{\mathbf{v}} = 1$ for $\mathbf{v} \in \xi_d$ such that $\sum v_i = 1$, and $\lambda_{\mathbf{v}} = 0$ for all other $\mathbf{v} \neq \mathbf{0}$. In this case

$$\epsilon_i(\mathbf{x}_{-i}) = 1 + \delta \prod_{j:j \neq i} x_j \geq 1. \quad (1.25)$$

Consider a Gibbs sampler step under the rule (1.23) for all points in the domain of (1.24); from (1.25), F_{ϵ} will need to apply to all ϵ in $E = [1, \infty)$. Yet, $\inf_{\epsilon \in E} F_{\epsilon} = \lim_{\epsilon \rightarrow \infty} F_{\epsilon} = 0$ and $\sup_{\epsilon \in E} F_{\epsilon} = F_1$ so after one full Gibbs sweep the upper bounding path initiated at $\{+\infty\}^d$ will have finite co-ordinates; the rest of the Gibbs steps will be carried out in the regular way. It can be shown that the bounding paths updated by the supremum and infimum of the maps F_{ϵ} converge.

1.7 Simulation results

For this section we use the abbreviations RS:=Rejection Sampling, PA:=Pure Algorithm and CSA:=Case-Specific Algorithm.

1.7.1 Running the algorithms

To try to eliminate the possibility of coding error, we first applied PA and CSA to the tri-variate distribution:

$$TN_3(\mathbf{0}, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}}, \quad \text{for } \mathbf{a} = \{-\infty\}^3, \mathbf{b} = \{+\infty\}^3, \quad (1.26)$$

$$Q = (q_{ij}) \text{ with } q_{ii} = 1, q_{ij} = -0.4 \text{ if } i \neq j,$$

and got two samples each of size 50,000. We used 9 Gibbs steps within each block for PA and 5 Gibbs steps for CSA before the attempt for coalescence. The marginal distribution for each co-ordinate of the target distribution is just $\mathcal{N}(0, 2.1426)$. The left qq-plot in Fig.1.7 and the Kolmogorov-Smirnov (KS) tests (Table 1.1) on all univariate samples are reassuring as to the correctness of our algorithms.

Co-ordinate	1st	2nd	3rd
PA	0.11	0.40	0.95
CSA	0.35	0.65	0.31

Table 1.1: p-values for the KS test on two samples each of size 50,000 from (1.26) after applying PA and CSA. The test was performed separately for each of the three co-ordinates. The null hypothesis in all cases is that the draws are from $\mathcal{N}(0, 2.1426)$.

To make sure that the algorithms are correct even when the state space is truncated we sample 50,000 draws from the distribution:

$$TN_3(\mathbf{0}, Q^{-1})|_{\mathbf{a}}^{\mathbf{b}}, \text{ for } \mathbf{a} = (0, 0, 0), \mathbf{b} = (10, 10, 10), \quad (1.27)$$

$$Q = (q_{ij}) \text{ with } q_{ii} = 1, q_{ij} = -0.4 \text{ if } i \neq j,$$

(5 and 3 Gibbs steps for PA and CSA respectively) and carry out similar error-checking tests, see the right qq-plot of Fig.1.7 and Table 1.7.1. This time the marginal distributions for each co-ordinate do not have a simple form so the results are compared with draws obtained from an RS that uses the corresponding untruncated density as envelope function. Sampling from unconstrained multivariate gaussian densities is carried out via the Cholesky decomposition of the variance-covariance matrix, see for instance ch.XI of Devroye (1986).

We repeated the same testing procedure for different dimensions and truncation regions (chosen in a way that the benchmark rejection sampling is not too inefficient) and did not inspect any flaws.

co-ordinate	PA-RS	PA-CSA	CSA-RS
1st	0.58	0.67	0.67
2nd	0.80	0.97	0.44
3rd	0.75	0.55	0.27

Table 1.2: p-values for the KS test on three samples each of size 50,000 from (1.27) after applying PA, CSA and RS. The test was performed separately for each of the three co-ordinates and for all possible pairs of algorithms. The null hypothesis in all cases is that the observations are from the same distribution.

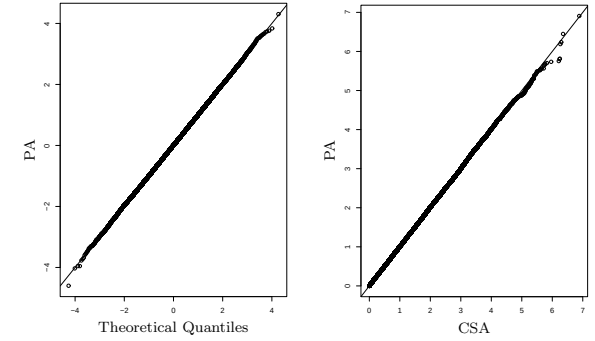


Figure 1.7: On the left, a qq-plot for the first co-ordinate of a sample of size 50,000 after implementing PA on (1.26). The observations have been divided with $\sqrt{2.1426}$ so that the estimated quantiles are comparable with those of the $\mathcal{N}(0, 1)$. On the right, a qq-plot comparing the quantiles for the distribution of the first co-ordinate of a random vector distributed according to (1.27) as estimated from two samples of size 50,000 from PA and CSA.

1.7.2 Efficiency evaluation

We compare PA and CSA with RS that proposes from the corresponding untruncated density. We consider a best case scenario for RS with $\mathbf{a} = \{0\}^d$ and $\mathbf{b} = \{10\}^d$; a big proportion of the probability mass for the envelope density of RS is in the truncation region so the probability that a draw from this envelope density gets accepted is big compared to other cases. It is interesting to incorporate in this comparison different positive correlation levels. An easy way to

		dimensionality			
		10	25	75	100
correlation level	RS				
	PA				
	CSA				
	0.2	362 7,712 (1) 6,671 (1)	0 2,421 (2) 2,139 (1)	0 435 (3) 508 (2)	0 264 (3) 492 (2)
	0.5	1,173 4,519 (3) 4,563 (2)	0.22 1,221 (5) 1,453 (3)	0 224 (7) 337 (3)	0 128 (8) 208 (4)
	0.8	5,882 1,266 (11) 2,325 (6)	27 154 (17) 667 (7)	0 4 (50) 167 (8)	0 1 (200) 120 (8)

Table 1.3: Draws/sec for RS that proposes from the corresponding unconstrained density, PA and CSA when used on $TN_d(\mathbf{0}, Q^{-1})|_{\mathbf{b}}^{\mathbf{a}}$ for $\mathbf{a} = \{0\}^d$ and $\mathbf{b} = \{10\}^d$ for different d 's and levels of positive correlation. In parentheses the number of Gibbs steps within a block for PA and CSA.

achieve that is by considering Stieltjes precision matrices $Q = (q_{ij})$ with $q_{ii} = 1$ and $q_{ij} = -c/(d-1)$ for some $0 < c < 1$ when $i \neq j$. This construction leads to appropriate symmetric, positive definite matrices. As c increases higher positive correlation levels are explored.

Table 1.3 shows the number of draws/sec after running the algorithms for different dimensions and different levels of positive correlation. The column of the positive correlations shows the value of c considered when constructing the precision matrix with the method described above. The table also gives in parentheses the number of Gibbs steps within a block (excluding the final coalescence Gibbs step) for each case of the PA and CSA. The number of Gibbs steps chosen for each algorithm is the one that maximised its efficiency in terms of the computational time to obtain a given number of samples. All the algorithms are written in C and were executed on an Athlon personal computer, running at 1500MHz in a Linux environment.

The cases considered in Table 1.3 are convenient not only for RS but also for CSA. PA can outperform CSA when truncations at more heterogeneous re-

gions are considered. For example, we applied PA and CSA to the 50-dimensional $TN_{50}(\mathbf{0}, Q^{-1})|_{\mathbf{b}}^{\mathbf{a}}$ for $\mathbf{a} = \{-40\}^{25} \times \{40\}^{25}$ and $\mathbf{b} = \{-20\}^{25} \times \{60\}^{25}$ and $Q = (q_{ij})$ with $q_{ii} = 1$ and $q_{ij} = -0.8/49$ for $i \neq j$. CSA is then inefficient because at the first Gibbs steps values $\Phi(x)$ from the cdf of the $\mathcal{N}(0, 1)$ distribution for $|x| > 38$ are usually requested (see the discussion at the end of Section 1.4). PA returns approximately 50 draws/sec and it needs only 7 Gibbs steps to deliver blocks with probability of being successful close to 0.8. The small number of Gibbs steps needed is due to the fact that for truncations distanced from the origin a huge proportion of the probability mass of the target gaussian density is found near the boundary closest to the origin, so the Gibbs sampler is rapidly convergent.

1.8 Conclusions

We have introduced a flexible CFTP methodology for MCMC algorithms. We also devised a powerful coupling technique (the Monotone Independence Coupler) which can be applied in various settings. The algorithms we presented use the Ro-CFTP protocol and rely on the construction of blocks consisting of three different stages (forcing compactness, bridging the gap, and attempting for one-shot coalescence). The construction of our coupling block allows many options for extension and variation which might be appropriate in different examples. We have presented our methods in a fairly general way, and though applied only to the TN family of target densities, we hope we have managed to illustrate the general applicability of the techniques introduced.

We have found that implementation on Gibbs samplers with dimension up to 100 is well within the feasibility of the method, though we have not as yet carried out more extensive simulation studies on distributions obtained (for instance) as posterior distributions in Bayesian analysis.

The Read-Once CFTP idea circumvents the need for storing random seeds, and

thus programming is generally easier and less prone to errors. The One-Shot idea spares time from attempting for coalescence when its probability is small. Note that the coalescence step, need not be monotone; assume that at the final step of a block of a Ro-CFTP algorithm it is possible that we use the splitting technique, i.e. the kernel $P(x, dy)$ under consideration can split as $\rho \nu(dy) + (1 - \rho) Q(x, dy)$ for any x between some appropriately constructed bounding paths. With probability ρ all paths are projected onto a draw from ν and the block is successful whereas with probability $1 - \rho$ we just declare the block unsuccessful and simply update the coalescence path according to the residual kernel $Q(x, dy)$.

There is nothing in the general methodology which restricts us to the use of Gibbs samplers. Independence Couplers can be constructed in a natural way for general Metropolis-Hastings kernels (see Breyer and Roberts, 2001). We did however exploit, for the example with the TN s, a monotonicity property of the Gibbs sampler and this was critical for the efficiency of the bounding techniques. Ongoing work will investigate the scope of the method for more general cases.

Acknowledgements

The first author would like to acknowledge financial support from the Greek State Scholarship's Foundation.

APPENDIX

Proof of Theorem 1.1:

Recall that $\mathcal{P}(x, dy) = P[\Phi(x) \in dy]$. Let $\nu(dy) := P[\Phi(x) \in dy | D]$ independent of x . The kernel $\mathcal{P}$ can split as following:

$$\mathcal{P}(x, dy) = \rho \nu(dy) + (1 - \rho) Q(x, dy), \quad (1.28)$$

for $\rho = P[D]$ and $Q(x, dy) \equiv P[\Phi(x) \in dy | D^c]$ the residual kernel.

We will need some transition (probability) kernel operations; the product of two transition kernels P_1, P_2 is the transition kernel $P_1 P_2(x, dy) = \int_z P_1(x, dz) P_2(z, dy)$, the power of a transition kernel is defined inductively as $P^n = P^{n-1} P$ with $P^0(x, dy) \equiv \mathbb{I}_{\{x \in dy\}}$ and the product of a probability measure λ with a transition kernel P is the probability measure $\lambda P(dx) = \int_z \lambda(dz) P(z, dx)$.

Let $W^i = \Phi^{\tau_i-1} \circ \dots \circ \Phi^{\tau_0+1} \circ \Phi^{\tau_0}(S)$, $i \geq 1$, be the random variable that refers to the position of the coalescence path one block-step before the $(i+1)$ -th successful block and Δt_i be the number of (unsuccessful) blocks between this successful block and the immediately previously successful one. It is then true that:

$$\begin{aligned} P[W^i \in dw] &= \sum_{j=0}^{\infty} P[\Delta t_i = j] P[W^i \in dw | \Delta t_i = j] = \\ &= \sum_{j=0}^{\infty} (1 - \rho)^j \rho \nu Q^j(dw) \equiv \pi(dw). \end{aligned}$$

The last and critical part of the above calculations is obtained algebraically after using (1.28) and that $\pi \mathcal{P} = \pi$, see Breyer and Roberts (2001) for details. The independence among the draws follows from the *regeneration* of the coalescence path after successful blocks, i.e. tours of the coalescence path at the time instances between a successful block and one block before the next successful one are iid.

Proof of Theorem 1.2:

(i) It is just a case of the Independence Coupler.

(ii) Get $F_x \sim P(x, dy)$. Consider the sets $R_1 = \{x \in X : x \leq x^*, F_x < Y\}$, $R_2 = \{x \in X : x^* \leq x, F_x > Y\}$ and

$$A = X \cap (R_1 \cup R_2)^c = \{x \in X : x \leq x^*, F_x \geq Y\} \cup \{x \in X : x^* \leq x, F_x \leq Y\}.$$

Define also

$$G(x, Y) := \frac{p(x, Y) p(x^*, F_x)}{p(x, F_x) p(x^*, Y)}, \quad x \in X.$$

Choose any $x_1, x_2 \in X$ with $x_1 \leq x_2$. To simplify the notation we write F_1, F_2 instead of F_{x_1}, F_{x_2} respectively. We examine the following cases:

Case 1: $x_1, x_2 \in A$.

If $x \in A$ then $G(x, Y) = \frac{p(x^*, F_x)}{p(x, F_x)} / \frac{p(x^*, Y)}{p(x, Y)} \geq 1$ from property (ii) of $p(x, y)$ given in Proposition 1.1, so $C \circ F_1 = C \circ F_2 = Y$.

Case 2: $x_1, x_2 \in R_1$.

We again exploit the monotonicity property of $p(x, y)$. Recall that $F_1 \leq F_2 < Y$.

$$\begin{aligned} \frac{G(x_2, Y)}{G(x_1, Y)} &= \frac{\frac{p(x_2, Y) p(x^*, F_2)}{p(x_2, F_2)}}{\frac{p(x_1, Y) p(x^*, F_1)}{p(x_1, F_1)}} \geq \frac{\frac{p(x_2, F_2) p(x^*, F_2)}{p(x_2, F_2)}}{\frac{p(x_1, F_2) p(x^*, F_1)}{p(x_1, F_1)}} = \frac{p(x^*, F_2) p(x_1, F_1)}{p(x_1, F_2) p(x^*, F_1)} \geq \\ &\geq \frac{p(x^*, F_1) p(x_1, F_1)}{p(x_1, F_1) p(x^*, F_1)} = 1. \end{aligned}$$

Since $G(x_2, Y) \geq G(x_1, Y)$ it is clear from (1.6) that if x_1 is mapped to Y ($U \leq G(x_1, Y)$) so is x_2 . This excludes, from the definition of R_1 , the only possibility of getting $C \circ F_1 > C \circ F_2$.

Case 3: $x_1, x_2 \in R_2$.

Exactly as in the previous case:

$$\begin{aligned} \frac{G(x_2, Y)}{G(x_1, Y)} &= \frac{\frac{p(x_2, Y) p(x^*, F_2)}{p(x_2, F_2)}}{\frac{p(x_1, Y) p(x^*, F_1)}{p(x_1, F_1)}} \leq \frac{\frac{p(x_2, F_1) p(x^*, F_2)}{p(x_2, F_2)}}{\frac{p(x_1, F_1) p(x^*, F_1)}{p(x_1, F_1)}} = \frac{p(x_2, F_1) p(x^*, F_2)}{p(x^*, F_1) p(x_2, F_2)} \leq \\ &\leq \frac{p(x_2, F_2) p(x^*, F_2)}{p(x^*, F_2) p(x_2, F_2)} = 1. \end{aligned}$$

Once again, after considering the definition of R_2 and that if x_2 accepts Y so does x_1 we get that $C \circ F_1 \leq C \circ F_2$.

envelope	called on	$r_{\alpha, \beta}$	inf (attained at)
$e^{-\alpha x + a^2/2} \mathbb{I}_{[\alpha, \beta]}(x)$	if $\alpha \geq 1/2$	$\frac{\alpha \sqrt{2\pi}(\Phi(\beta) - \Phi(\alpha))}{\exp(-\alpha^2/2) - \exp(a^2/2 - \alpha\beta)}$	0.44^a ($\alpha, \beta = (1/2, \infty)$)
$e^{-\beta x + \beta^2/2} \mathbb{I}_{[\alpha, \beta]}(x)$	else if $\beta \leq -1/2$	$\frac{(-\beta) \sqrt{2\pi}(\Phi(\beta) - \Phi(\alpha))}{\exp(-\beta^2/2) - \exp(-\alpha\beta + \beta^2/2)}$	0.44^b ($\alpha, \beta = (-\infty, -1/2)$)
$\mathbb{I}_{[\alpha, \beta]}(x)$	else if $(\beta - \alpha) \leq 3/2$	$\sqrt{2\pi} \frac{\Phi(\beta) - \Phi(\alpha)}{\beta - \alpha}$	0.47^c ($\alpha, \beta = (1/2, 2)$)
$e^{-x^2/2}$	else	$\Phi(\beta) - \Phi(\alpha)$	0.29^d ($\alpha, \beta = (1/2, 2)$)

^a r is $\downarrow \beta$ (decreasing in β) for constant α and $\uparrow \alpha$ (increasing in α) for constant $\beta - \alpha$.

^b r is $\uparrow \alpha$ for constant β and $\downarrow \beta$ for constant $\beta - \alpha$.

^cFor constant $\beta - \alpha \equiv c$, r is $\uparrow \beta \in (-1/2, c/2]$ and $\downarrow \beta \in [c/2, 1/2 + c]$, so $r_{\alpha, \beta} \geq r_{1/2, 1/2+c}$ which is $\downarrow c$, so $r_{\alpha, \beta} \geq r_{1/2, 2}$.

^d $r_{\alpha, \beta} \geq r_{\alpha, \alpha+3/2}$ which is $\downarrow \alpha$.

Figure 1.8: The method for drawing from the univariate truncated gaussian density $TN_1(\mu, 1)|_\alpha^\beta$.

From the definition of the sets A, R_1, R_2 and the result for the case when $x \in A$ it is straightforward that the only case that $x_1 \leq x_2$ can lead to $C \circ F_1 > C \circ F_2$ is when $x_1, x_2 \in R_1$ or $x_1, x_2 \in R_2$. For instance, if $x_1 \in R_1$ and $x_2 \in R_2$ then clearly $C \circ F_1 \leq Y \leq C \circ F_2$, if $x_1 \in R_1$ and $x_2 \in A$ then $C \circ F_1 \leq Y = C \circ F_2$, and so on. The preceding results demonstrate that even when $x_1, x_2 \in R_1$ or $x_1, x_2 \in R_2$ monotonicity, i.e. $C \circ F_1 \leq C \circ F_2$ for $x_1 \leq x_2$, is preserved.

The algorithm that draws from $TN_1(\mu, 1)|_\alpha^\beta$:

The algorithm incorporates four separate rejection sampling algorithms. The acceptance probability $r_{\alpha, \beta}$ for each case and its minimum value are given in Fig.1.8 (to simplify the demonstration we have set $\mu = 0$ so the target density is proportional to $e^{-x^2/2} \mathbb{I}_{[\alpha, \beta]}(x)$).

The methodology is close to the one proposed in Geweke (1991) except for avoiding using the cdf of the $\mathcal{N}(0, 1)$ distribution when choosing among available comparison functions. That guarantees the unrestricted applicability of the resulted algorithm. In Robert (1995) a slightly different methodology is proposed which yields better acceptance probabilities for the proposed variates but as al-

ready noted in Geweke (1991) the increased algebraic structure of this algorithm renders it equivalent to ours in terms of computational time.

Proof of Proposition 1.2:

Certainly, $\lambda_V > 0$ since V is positive definite. Also, there exists an orthogonal matrix $T \in \mathbf{R}^{d \times d}$ ($T^\top T = I$) such that $T^\top \Lambda T = V$ where $\Lambda = (\lambda_{ij})$ is a diagonal matrix with diagonal elements the eigenvalues of V . Then, for $\mathbf{y} = T\mathbf{x}$:

$$\mathbf{x}^\top V \mathbf{x} = (T\mathbf{x})^\top \Lambda (T\mathbf{x}) = \sum_{i=1}^d \lambda_{ii} y_i^2 \geq \lambda_V \sum_{i=1}^d y_i^2 = \lambda_V \mathbf{y}^\top \mathbf{y} = \lambda_V \mathbf{x}^\top \mathbf{x} .$$

Article 2

Exact Simulation of Diffusions

2.1 Introduction

Exact simulation of SDEs is a notorious problem within the applied probability community. The objective of this paper is to present a first step towards the solution of this problem. We describe a new algorithm we call the Exact Algorithm for simulating a class of SDEs. It involves rejection sampling and, when applicable, returns *exact* draws from any finite-dimensional distribution of the solution of the SDE.

Let $\{W_t; t \in [0, T]\}$ be a scalar Brownian motion. Consider the general type of the one-dimensional Itô's diffusion:

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \in [0, T], \quad X_0 = x \in \mathbf{R}, \quad (2.1)$$

for drift coefficient $b : \mathbf{R} \mapsto \mathbf{R}$ and diffusion coefficient $\sigma : \mathbf{R} \mapsto \mathbf{R}$. Under regularity conditions on b and σ it can be shown that (2.1) has a solution $\{X_t; t \in [0, T]\}$ weakly unique, i.e. all the solutions have identical finite-dimensional distributions. Weak uniqueness, relatively more general than path-wise uniqueness, is sufficient for simulation purposes. For a formal definition of (2.1) see for instance McKean (1969).

Mathematical models of this kind are used to describe the evolution of stochastic phenomena in a wide range of disciplines and most times it is important to be able to simulate them. In rare cases (2.1) has an explicit solution with identifiable transition density. However, in most of the models used in practice it is necessary to resort to a numerical solution of (2.1). This has traditionally implied the use of some of the discrete time approximation methods (Euler, Taylor's expansion, etc.) which rely on approximate small time increment distributions for the diffusion (for a detailed account of these methods see Kloeden and Platen, 1992). In many cases throughout this paper the results of our algorithm will be compared with the Euler scheme which approximates (2.1) via the recursion:

$$X_{t+\Delta} = X_t + b(X_t)\Delta + \sigma(X_t) \cdot \mathcal{N}(0, \Delta) .$$

For a large class of processes the Exact Algorithm provides an alternative which involves no approximation (apart from that inherent in any computer simulation) and yet is computationally highly efficient. It returns skeletons of *exact* paths which can be easily *filled in* without further reference to the diffusion dynamics. Thus the method can be used to simulate the diffusion at a prescribed collection of time points, or alternatively at times which occur to be interesting to the user, *after* the completion of the algorithm.

We begin (Section 2.2) by stating the rejection sampling technique in a way that serves our purposes. In Section 2.3 we present our method and in Section 2.4 give some results related with its efficiency. In Section 2.5 we apply the algorithm to a specific SDE and in Section 2.6 we take advantage of the properties of the Exact Algorithm to simulate exactly extremes and hitting times for the same SDE. Finally (Section 2.7), we present some ideas that could, in future research, overcome the restrictions of the algorithm and give some general conclusions. We are going to restrict the exposition of our algorithm for the case when $\sigma = 1$. This

is not by any means restrictive since the SDE (2.1) can be transformed into one of unit diffusion coefficient for the process $Y = \{Y_t; t \in [0, T]\}$ defined as:

$$Y_t = \int_z^{X_t} \frac{1}{\sigma(u)} du ,$$

where z is an arbitrary element of the state space of X .

2.2 A general rejection sampling algorithm

Rejection Sampling (RS) is a widely used simulation technique. It is frequently presented as follows. Assume that f, g are probability densities w.r.t. some measure on $\mathbf{R}^d$ and that there exists $\epsilon > 0$ such that $\epsilon \frac{f}{g} \leq 1$. Then the iterative algorithm:

Rejection Sampling

1. **Sample** $Y \sim g$.
 2. **Sample** $U \sim \text{Un}(0, 1)$.
 - 3₁ **If** $U < \epsilon \frac{f}{g}(Y)$ **return** Y .
 - 3₂ **Else go to** (1).
-

returns an observation distributed according to f .

This traditional rejection sampling algorithm seems to imply a fixed order in the acquisition of the required random elements: specifically, the proposed variate Y precedes the decision variate U . However, such a prescribed order for the simulation steps is not necessary. The algorithm to be presented in Section 2.3 is more easily understood when imagining that the proposed variate actually *succeeds* the decision variate. Moreover, this reordering of the random steps can be carried out without adding any complexity to the algorithm. See Papaspiliopoulos and Roberts (2004) for a case where a similar reordering of the random inputs is used to perform an otherwise impossible MCMC algorithm for the Dirichlet mixture model.

Additionally, in some cases there are ways of constructing a condition for the acceptance or the rejection of the current proposed element Y from minimal information about it. This will be essential in the diffusion context where it will never be possible to store a complete, continuous path. A similar idea appears in Kendall (1997) on a perfect simulation algorithm of point processes. In that case a probability related with a complicated surface is expressed as the probability of an appropriately constructed event whose truth or otherwise is easy to verify.

We now present a formal definition of the RS algorithm in a way that incorporates the observations just described. Let $(C, \mathcal{C})$ be a sufficiently regular measurable space and ν, μ probability measures on it such that μ is absolutely continuous w.r.t. ν . Assume that there exists $\epsilon > 0$ such that $f := \epsilon \frac{d\mu}{d\nu} \leq 1$ ν -a.s. and that it is easy to sample from ν . The following proposition can be used to return draws from μ .

Proposition 2.1 (Rejection Sampling) *Let $(Y_n, I_n)_{n \geq 1}$ be a sequence of iid random elements in $C \times \{0, 1\}$ such that $Y_1 \sim \nu$ and $P[I_1 = 1 | Y_1 = y] = f(y)$ for all $y \in C$. Define $\tau = \min\{i \geq 1 : I_i = 1\}$. Then $P[Y_\tau \in dy] = \mu(dy)$.*

Proof: In the Appendix.

This presentation of the RS scheme does not assume any order for the simulation of Y and I and, besides the certain conditional property given in the Proposition, does not restrict in any other way the construction of the binary indicator I .

2.3 The Exact Algorithm

Consider the stochastic process $X = \{X_t; t \in [0, T]\}$ determined as the unique solution of the SDE:

$$dX_t = \alpha(X_t) dt + dW_t, \quad t \in [0, T], \quad X_0 = 0. \quad (2.2)$$

The drift function $\alpha : \mathbf{R} \mapsto \mathbf{R}$ is presumed to satisfy the regularity properties that guarantee the existence of a global, weakly unique solution for (2.2). In particular, it suffices that α is locally Lipschitz, i.e. for each $M > 0$ there exists $K_M > 0$ such that:

$$|\alpha(y) - \alpha(x)| \leq K_M |y - x|, \quad \text{for all } |x| \leq M, |y| \leq M,$$

with a linear growth bound; there exists $K > 0$ such that

$$|\alpha(x)|^2 \leq K^2(1 + x^2), \quad x \in \mathbf{R}.$$

See ch.4 of Kloeden and Platen (1992) for a detailed presentation of weaker conditions.

Since we are going to carry out rejection sampling it is convenient to think of the stochastic processes involved as measures induced on the space C of continuous functions from $[0, T]$ to $\mathbf{R}$. We denote by $\omega = \{\omega_t; t \in [0, T]\}$ a typical element of this space and by $\mathcal{C}$ the cylinder σ -algebra on C generated by the co-ordinate mappings. The process $W = \{W_t; t \in [0, T]\}$ is a Brownian motion started at 0 and $\mathbb{W}$ the Wiener measure on $(C, \mathcal{C})$ corresponding to its law.

2.3.1 Rejection sampling for SDEs

We explain how a rejection sampling algorithm can be set up for the case of the SDE given in (2.2). The analysis begins with the Girsanov transformation of measures. Let $\mathbb{Q}$ be the probability measure on $(C, \mathcal{C})$ corresponding to the law of the target process $X = \{X_t; t \in [0, T]\}$.

Proposition 2.2 (Girsanov Transformation) *Assume that the drift coefficient α satisfies Novikov's condition:*

$$\mathbb{E}_{\mathbb{W}} \left[\exp \left\{ \frac{1}{2} \int_0^T \alpha^2(\omega_t) dt \right\} \right] < \infty.$$

It is then true that:

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(\omega) = \exp \left\{ \int_0^T \alpha(\omega_t) d\omega_t - \frac{1}{2} \int_0^T \alpha^2(\omega_t) dt \right\} =: G(\omega) . \quad (2.3)$$

Proof: See for instance Oksendal (1998), ch.8.

Our goal is to implement a rejection sampling algorithm using (2.3) to construct an accept-reject mechanism. The difficulty is that exact evaluation of $G(\omega)$ is impossible. We can simplify $G(\omega)$ using Itô's formula to remove the Itô's integral term and we shall see that this allows us to carry out the rejection scheme indirectly. We now need our first assumption.

CONDITION 1:

The drift coefficient α is continuously differentiable.

Under Condition 1, $G(\omega)$ admits the following simplification; let $A(u) = \int_0^u \alpha(y) dy$, $u \in \mathbf{R}$. Itô's formula then gives that:

$$\int_0^T \alpha(\omega_t) d\omega_t = A(\omega_T) - A(\omega_0) - \frac{1}{2} \int_0^T \alpha'(\omega_t) dt .$$

We can now write $G(\omega)$ as:

$$G(\omega) = \exp \left\{ A(\omega_T) - A(\omega_0) - \frac{1}{2} \int_0^T \left(\alpha^2(\omega_t) + \alpha'(\omega_t) \right) dt \right\} .$$

Rejection sampling using Brownian candidates is only conceivably possible if $G(\omega)$ is almost surely bounded and this is likely to require A to be bounded. To remove this requirement we introduce a third probability measure which will be used to construct the candidates for the rejection sampling scheme.

Consider the *biased* Brownian motion heuristically defined as $(W | W_T = \rho)$ with ρ distributed according to some density function $h : \mathbf{R} \mapsto [0, \infty)$ w.r.t. the Lebesgue measure. Let $\mathbb{Z}$ be the law of the biased Brownian motion.

Proposition 2.3 (Biased Brownian Motion) *If the support of h is the real line, then $\mathbb{Z}$ is equivalent to $\mathbb{W}$ and:*

$$\frac{d\mathbb{Z}}{d\mathbb{W}}(\omega) = \frac{h(\omega_T)}{\frac{1}{\sqrt{2\pi T}} \exp \left(-\frac{\omega_T^2}{2T} \right)} .$$

Proof: In the Appendix.

It is now trivial that:

$$\frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) = \frac{d\mathbb{Q}}{d\mathbb{W}} \frac{d\mathbb{W}}{d\mathbb{Z}}(\omega) \propto \exp \left\{ A(\omega_T) - \frac{\omega_T^2}{2T} - \frac{1}{2} \int_0^T \left(\alpha^2(\omega_t) + \alpha'(\omega_t) \right) dt \right\} / h(\omega_T) ,$$

where $\propto$ implies that we omitted some factors not depending on ω . Under the condition that:

CONDITION 2:

$$\int_{\mathbf{R}} \exp \{ A(u) - u^2/2T \} du =: c < \infty$$

and after choosing $h(u) = \exp \{ A(u) - u^2/2T \} / c$,

$$\frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) \propto \exp \left\{ - \int_0^T \left(\frac{1}{2} \alpha^2(\omega_t) + \frac{1}{2} \alpha'(\omega_t) \right) dt \right\} . \quad (2.4)$$

Assume now that the functional involved in the above integral is bounded:

CONDITION 3:

There exist constants $k_1, k_2 \in \mathbf{R}$ such that $k_1 \leq \frac{1}{2} \alpha^2(u) + \frac{1}{2} \alpha'(u) \leq k_2$, $u \in \mathbf{R}$.

We can then write (2.4) as:

$$\frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) \propto \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\}, \quad (2.5)$$

for a function $\phi \geq 0$ defined as $\phi(u) = \frac{1}{2}\alpha^2(u) + \frac{1}{2}\alpha'(u) - k_1$, $u \in \mathbf{R}$. This creates the possibility of performing rejection sampling with candidates from $\mathbb{Z}$ in order to sample from $\mathbb{Q}$ which is the objective. We can choose the length $T > 0$ of the time interval under consideration so that:

$$0 \leq \phi(u) \leq T^{-1} \text{ for any } u \in \mathbf{R}. \quad (2.6)$$

Just consider any $T \leq 1/(k_2 - k_1) = 1/R$ where R is the *identifiable* range of $(\alpha^2 + \alpha')/2$ in the sense that the lower (upper) bound we can actually obtain for $(\alpha^2 + \alpha')/2$ can be less (greater) than its maximum (minimum) value. From this point on it is assumed that T has been fixed so that (2.6) is true.

We remark that Condition 3 is trivially implied by

CONDITION 3':

There exist constants $k'_1, k'_2 \in \mathbf{R}$ such that $k'_1 \leq \alpha(u), \alpha'(u) \leq k'_2$, $u \in \mathbf{R}$.

Furthermore, Condition 3' implies that Condition 2 holds and moreover the constants k'_1 and k'_2 point the way to implementing simple rejection sampling algorithms for simulating the end-point of the biased Brownian motion.

2.3.2 Constructing the Exact Algorithm

We set $H(\omega) = \int_0^T \phi(\omega_t) dt$. We assume from now on that we have access to paths $\omega \sim \mathbb{Z}$ of the biased Brownian motion. Note that we can realise such a path at any finite collection of time instances by drawing first its ending point $\omega_T \sim h$ and then the rest of the skeleton according to the dynamics of a Brownian bridge. Drawing

from the univariate distribution h cannot be a big problem; Devroye (1986) gives many algorithms for drawing from densities on $\mathbf{R}$ that could be used as envelopes for a rejection sampling scheme on h .

The preliminaries of Section 2.3.1 together with the general rejection sampling protocol of Proposition 2.1 ensure that the following algorithm (were it implementable in practice) would output realisations of the diffusion X that solves the SDE (2.2):

Impossible Algorithm

1. **Sample a complete, continuous path** $\omega \sim \mathbb{Z}$.
 2. **Compute** $H(\omega)$.
 3. **Produce a binary variable** I **so that** $P[I = 1 | \omega] = \exp\{-H(\omega)\}$.
 4. **If** $I = 0$ **go to** (1).
 5. **Output** ω .
-

The indicator I is easily constructed with the use of some $U \sim \text{Un}(0,1)$. In practise, we can only simulate the path ω at any given finite collection of instances $0 \leq t_1, t_2, \dots, t_n \leq T$ so evaluating the integral $H(\omega)$ is impossible. However, given that Conditions 1–3 are satisfied, we can produce an algorithm which manages to circumvent Steps 1 and 2 and still carry out Steps 3 and 4 exactly given only a *finite* but random skeleton of instances of ω .

The idea builds on the simple observation that for the bounded continuous function $0 \leq \phi(\omega_t) \leq T^{-1}$ events of probability $\int_0^T \phi(\omega_t) dt$ can be constructed simply by drawing a random point $(V, W) \sim \text{Un}[(0, T) \times (0, T^{-1})]$. Then, the event $\{\phi(\omega_V) \geq W\}$ will have the required probability.

To extend this idea to an event of probability $\exp\{-H(\omega)\}$ we exploit a Taylor series expansion construction which gives us an event of probability $\exp\{-H(\omega)\}$ as an event which depends upon a countable sequence of events of probability $H(\omega)$. Furthermore, it turns out to be possible to express this event both as the countable union of a sequence of increasing events and the countable intersection of another sequence of decreasing events. Crucially, for each event in either of the two sequences its truth or otherwise can be confirmed by a finite skeleton of a path $\omega \sim \mathbb{Z}$ and, consequently, the truth or otherwise of the event of probability $\exp\{-H(\omega)\}$ can also be determined after finite computations.

All the above ideas are presented in a rigorous way in Theorem 2.1 that follows. The construction to be described is similar in spirit, though in a different context, with Von Neumann's comparison method for the simulation of exponential random variables, see Forsythe (1972) for a detailed review. We have denoted by $(\Omega, \mathcal{F}, P)$ the underlying probability space that generates all the random elements involved in the theorem.

Theorem 2.1 *Let $\omega \sim \mathbb{Z}$ be a path of the biased Brownian motion on $[0, T]$. Let $\tau = (V_i, W_i)_{i \geq 1}$ be a sequence of iid points uniformly distributed on $(0, T) \times (0, 1/T)$. Consider also some $U \sim \text{Un}(0, 1)$. Assume that ω , τ , and U are independent. Define the following events:*

$$\Gamma_0 = \Omega, \quad \Gamma_n = \left\{ \phi(\omega_{V_1}) \geq W_1, \dots, \phi(\omega_{V_n}) \geq W_n, U \leq \frac{1}{n!} \right\}, \quad (2.7)$$

for $n \geq 1$. Consider the sequence of events $(E_n)_{n \geq 1}$ defined as:

$$\begin{aligned} E_{2n+1} &= (\Gamma_0 - \Gamma_1) + (\Gamma_2 - \Gamma_3) + \dots + (\Gamma_{2n} - \Gamma_{2n+1}), \\ E_{2n+2} &= (\Gamma_0 - \Gamma_1) + (\Gamma_2 - \Gamma_3) + \dots + (\Gamma_{2n} - \Gamma_{2n+1}) + \Gamma_{2n+2}, \end{aligned} \quad (2.8)$$

$n \geq 0$, where $(+)$ implies union of disjoint sets and $D - F = D \cap F^c$ for any sets $F \subseteq D$. Then:

- (i) $(E_{2n+1})_{n \geq 0}, (E_{2n+2})_{n \geq 0}$ are sequences of increasing and decreasing events respectively with $E_{2\kappa+1} \subseteq E_{2\lambda+2}$ for any $\kappa, \lambda \in \{0, 1, \dots\}$ with the limiting property that $P[\bigcap_0^\infty E_{2n+2} - \bigcup_0^\infty E_{2n+1}] = 0$.
- (ii) Let $E = \bigcup_0^\infty E_{2n+1}$. If I is a binary indicator such that $I = 1$ when E occurs and 0 otherwise then:

$$P[I = 1 | \omega] = \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\}.$$

Proof: In the Appendix.

In practice, we can simply identify $\bigcap_0^\infty E_{2n+2} - \bigcup_0^\infty E_{2n+1}$ with the null event

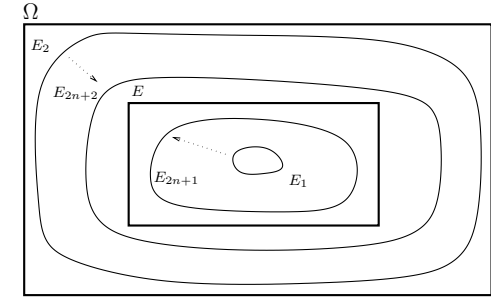


Figure 2.1: An illustration of the sets E_i , $i \geq 1$. When some even-numbered E_i does not occur or an odd-numbered E_i occurs, E does not occur or occurs respectively.

$\emptyset$ so that $E = \bigcup_0^\infty E_{2n+1} = \bigcap_0^\infty E_{2n+2}$. Fig.2.1 illustrates the decomposition of Ω over the sets E_i , $i \geq 1$.

Theorem 2.1 does not impose any restriction on the order of the realisation of the random elements when rejection sampling will take place. To transform the theorem into a feasible rejection sampling algorithm it is necessary that we draw U and then generate the path $\omega \sim \mathbb{Z}$ and the sequence τ in parallel. We follow an iterative process that involves drawing (V_i, W_i) uniformly from $(0, T) \times (0, 1/T)$

and then simulating ω_{V_i} conditionally on the already obtained $\omega_{V_1}, \omega_{V_2}, \dots, \omega_{V_{i-1}}$, for $i \geq 1$. Recall that this recursive construction of the biased Brownian motion path is straightforward as long as someone begins by simulating its ending point $\omega_T \sim h$ (the choice for h is given after Condition 2 of Section 2.3.1). Then, the rest of the path is a Brownian bridge; given that the locations $\{\omega_{V_1}, \dots, \omega_{V_{i-1}}, \omega_T\}$ have been constructed the path can be realised at the instance V_i just by drawing:

$$\omega_{V_i} \sim \mathcal{N}\left(\omega_{V_-} + \frac{\omega_{V_+} - \omega_{V_-}}{V_+ - V_-}(V_i - V_-), \frac{(V_+ - V_i)(V_i - V_-)}{V_+ - V_-}\right),$$

where we have defined the instances $V_- = \max\{0, V_j, j = 1, \dots, i-1 : V_j < V_i\}$ and $V_+ = \min\{T, V_j, j = 1, \dots, i-1 : V_j > V_i\}$. See for instance pg.360 of Karatzas and Shreve (1991) for the representation of a Brownian bridge as a transformation of an unconditional Brownian motion which implies the above formula.

From the definition of $(E_i)_{i \geq 1}$ it is clear that after j iterations are carried out we have the necessary information to decide if any of the events $E_1, E_2, \dots, E_j$ occurred or not. We perform iterations until the first time that an odd-numbered E_i occurs or an even-numbered E_i does not occur. It is then clear from Theorem 2.1 that the former case gives $I = 1$ and the latter $I = 0$. Further realisation of the random elements will not change the decision about I . If needed, we can continue simulating an accepted path of $\mathbb{Z}$ so that the resulted path of X is constructed at any time instances requested.

The recursive definition of $(E_i)_{i \geq 1}$ allows for a simple way of carrying out the successive steps of the iterative method described above. Assume for instance that $2n$ iterations have taken place without a decision about I , that is E_{2n} happened and E_{2n-1} did not happen. From (2.7), (2.8) it is clear that:

$$E_{2n} = E_{2n-1} + \Gamma_{2n} \quad \text{and} \\ E_{2n+1} = E_{2n} - \Gamma_{2n+1} = E_{2n-1} + (\Gamma_{2n} - \Gamma_{2n+1}).$$

The first equation indicates that Γ_{2n} occurred and the second that we only need to check if the sub-event of Γ_{2n}

$$\Gamma_{2n} - \Gamma_{2n+1} = \Gamma_{2n} \cap \left\{ \phi(\omega_{V_{2n+1}}) < W_{2n+1} \text{ or } U > \frac{1}{(2n+1)!} \right\}$$

occurred (i.e. $\phi(\omega_{V_{2n+1}}) < W_{2n+1}$ or $U > \frac{1}{(2n+1)!}$) or not to reach to a similar conclusion about E_{2n+1} . The same convenient interpretation appears for the case when an even-numbered iteration, say the $2n$ -th, is carried out. Given that I is not determined before that step we can conclude that E_{2n} did not take place if $U > 1/(2n)!$ or $\phi(\omega_{V_{2n}}) < W_{2n}$.

We can now present the pseudo-code that implements Theorem 2.1 to carry out rejection sampling:

Exact Algorithm

1. **Initiate a path of $\mathbb{Z}$:** Set $\omega_0 = 0$ and draw $\omega_T \sim h$.
 2. Draw $U \sim \text{Un}(0, 1)$. Set $i = 0$.
 3. Draw $(V, W) \sim \text{Un}[(0, T) \times (0, 1/T)]$. Set $i = i + 1$.
 4. **Construct ω_V given the currently simulated instances of ω .**
 - 5₁ **If $\phi(\omega_V) < W$ or $U > 1/i!$ then**
 - If i is even set $I = 0$ and go to (1).**
 - If i is odd set $I = 1$ and go to (6).**
 - 5₂ **Else go to (3).**
 6. **Output the currently simulated instances of ω .**
-

As already mentioned, we can intervene a step that constructs the proposed path at any time instances we require. It is now clear that the algorithm returns exact skeletons of the target process X at any given finite collection of time instances.

2.3.3 A factorisation for $\mathbb{Q}$

From a probabilistic point of view, the Exact Algorithm manages to decompose the target measure $\mathbb{Q}$ in terms of a product of Brownian bridges after appropriately extending the underlying probability space.

Analytically, let $S = \{(0, 0), (u_1, y_1), \dots, (u_k, y_k), (T, y_{k+1})\}$ be the random skeleton produced by the Exact Algorithm, for some random integer $k > 1$. If $\mathcal{S}_n$ is the space of the possible configurations of n points in $[0, T] \times \mathbf{R}$, $n \geq 1$, then S takes values in $\cup_n \mathcal{S}_n$. We avoid details about the σ -algebra construction on $\cup_n \mathcal{S}_n$, and simply denote by $\mathcal{L}_S$ the distribution of S . Let $\mathbb{B}\mathbb{B}(s, t; x, y)$, for $s \leq t$, and $x, y \in \mathbf{R}$, be the probability measure corresponding to a Brownian bridge starting at the time instance s from x and finishing at the time instance t at y . In terms of probability measures, the Exact Algorithm manages to factorise $\mathbb{Q}$ in the following way:

$$\mathbb{Q} = \bigotimes_{i=1}^{k+1} \mathbb{B}\mathbb{B}(u_{i-1}, u_i; y_{i-1}, y_i) \otimes \mathcal{L}_S(dS), \quad (2.9)$$

where $(u_0, y_0) \equiv (0, 0)$ and $u_{k+1} = T$. Critically, the rejection sampling construction of the Exact Algorithm allows for the simulation of $\mathcal{L}_S$, so drawing from $\mathbb{Q}$ is then straightforward. Fig.2.2 gives a graphical illustration of the factorisation (2.9).

Because of this decomposition of $\mathbb{Q}$ it is possible to identify characteristics of the target process X after considering the properties of the Brownian bridges that fill in its skeletons. Thus, it is possible to simulate exactly hitting times, extremes (we present these applications analytically in Section 2.6) and any other random elements for which there are explicit results for the case of Brownian paths.

Another simple example when we can exploit (2.9) is at the Monte Carlo evaluation of the expected value of functionals of X_t for some time instance $t \in [0, T]$. Assume that $\{X_t^i, S^i\}_{1 \leq i \leq n}$ are n draws for the skeleton S and X_t produced by

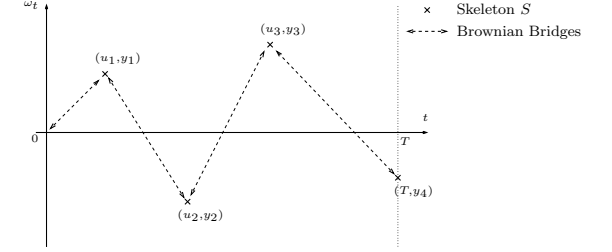


Figure 2.2: The factorisation of $\mathbb{Q}$: drawing from $\mathbb{Q}$ is achieved after simulating the skeleton $S = \{(0, 0), (u_1, y_1), \dots, (u_k, y_k), (T, y_{k+1})\}$ (in the case of the figure, $k = 3$) and then filling in the rest of the path with independent Brownian bridges.

the Exact Algorithm. An estimator of $\mathbb{E}[f(X_t)]$, for some function f , can be the mean of the $f(X_t^i)$, $1 \leq i \leq n$. Using the simple conditional expectation property $\text{Var}[f(X_t)] \geq \text{Var}[\mathbb{E}[f(X_t) | S]]$ we can produce an unbiased estimator of smaller variance after considering the mean of the $\mathbb{E}[f(X_t) | S^i]$, $1 \leq i \leq n$. Conditionally on the skeleton, X_t follows a normal distribution, so for reasonable functions f finding the $\mathbb{E}[f(X_t) | S]$ will be straightforward.

2.4 Efficiency of the Exact Algorithm

Two aspects of the Exact Algorithm need to be examined. The first involves the probability of accepting a proposed path from $\mathbb{Z}$ as a path of X . The second has to do with the number of points in $(0, T) \times (0, 1/T)$ required to reach to a decision about a proposed path.

We can rewrite (2.5) as:

$$\epsilon(T) \frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) = \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\}$$

for some appropriate $\epsilon(T)$. Note that $\mathbb{Q}$ and $\mathbb{Z}$ are both probability measures so $\epsilon(T)$ equals precisely the probability that a path $\omega \sim \mathbb{Z}$ is accepted as a path from $\mathbb{Q}$. Equivalently, $\epsilon(T) = \mathbb{P}[I = 1]$ for the binary indicator I defined in Theorem

2.1.

Proposition 2.4 *For any appropriate T , the probability $\epsilon(T)$ of the event $\{I = 1\}$ for I defined in Theorem 2.1 is at least e^{-1} and is decreasing in T with $\lim_{T \downarrow 0} \epsilon(T) = 1$.*

Proof: In the Appendix.

As expected, the probability of accepting the proposed paths increases when we decrease the length of the time period under consideration.

It is important that we give a result for the number of points uniformly drawn in $(0, T) \times (0, 1/T)$ required to accept or reject a proposed path. Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space that generates all the random elements involved in Theorem 2.1.

Proposition 2.5 *Define the random variable $N : \Omega \mapsto \{1, 2, \dots\}$ as:*

$$N(w) = \begin{cases} \min \{n = 2, 4, \dots : w \notin E_n\} & \text{if } w \in E^c \\ \min \{n = 1, 3, \dots : w \in E_n\} & \text{if } w \in E \end{cases}.$$

Then $\mathbb{E}[N] \leq e$.

Proof: Just note that $\mathbb{P}[N \geq n] \leq \mathbb{P}[U \leq \frac{1}{(n-1)!}]$, so it is straightforward that

$$\mathbb{E}[N] \leq \sum_{n=1}^{\infty} \frac{1}{(n-1)!} = e.$$

□

The random variable N counts the number of events E_n we have to consider before deciding if w belongs to E or E^c (equivalently, if the corresponding realisations of the random elements involved in Theorem 2.1 make E happen or not). It is a perhaps surprising result that for any eligible T and any drift coefficient α we are expecting on average less than three of the points drawn uniformly from

$(0, T) \times (0, 1/T)$ before we decide for the acceptance or the rejection of a proposed path.

Note that the Exact Algorithm can take advantage of the Markov property of the process X and produce skeletons of any requested length $l > 0$ after merging skeletons of lengths acceptable by the algorithm. The choice of the length of the merged skeletons and, subsequently, the efficiency of the Exact Algorithm depends on the identifiable range of the functional $(\alpha^2 + \alpha')/2$ of the drift function. Recall that the identifiable range is a constant R we can analytically find which satisfies

$$\sup_{x \in \mathbf{R}} (\alpha^2 + \alpha')(x)/2 - \inf_{x \in \mathbf{R}} (\alpha^2 + \alpha')(x)/2 \leq R.$$

The following proposition gives a result for the case when the Exact Algorithm merges skeletons of the maximum eligible length $T = 1/R$ to obtain a skeleton of length l .

Proposition 2.6 *If R is the identifiable range of the functional $(\alpha^2 + \alpha')/2$ of the drift α and N_l the total number of the uniformly drawn points needed for the Exact Algorithm to return a skeleton of length $l > 0$ then:*

$$\mathbb{E}[N_l] \leq [l \cdot R] \times e^2.$$

Proof: In the Appendix.

In total, the Exact Algorithm requires elementary computational skills. Except for the appealing characteristic of being exact it seems to compete with conventional approximation techniques even in terms of time efficiency. The example that follows favours this assertion.

2.5 Applying the Exact Algorithm

We apply the Exact Algorithm to the SDE:

$$dX_t = \sin(X_t) dt + dW_t, \quad t \in [0, T], \quad X_0 = 0. \quad (2.10)$$

Conventional methods can only approximate sample paths for the solution X of (2.10) after resorting to one of the suggested time discretisation techniques; the Exact Algorithm returns exact skeletons of X .

The drift coefficient $\alpha \equiv \sin$ satisfies Conditions 1 and 2. It is also easy to check that $-1/2 \leq \frac{1}{2} \sin^2(u) + \frac{1}{2} \cos(u) \leq 5/8$ for any $u \in \mathbf{R}$, so Condition 3 is satisfied for $k_1 = -1/2$ and $k_2 = 5/8$. In the present context $\phi(u) = \frac{1}{2} \sin^2(u) + \frac{1}{2} \cos(u) + \frac{1}{2}$. We can now choose $T = 1/(k_2 - k_1) = 8/9$. For this ending time instance, $0 \leq \phi(u) \leq T^{-1}$ for any real u . The biased process of law $\mathbb{Z}$ is defined as the conditioned Brownian motion $(W | W_T = \rho)$ with $\rho \sim h \propto \exp\{-\cos(u) - u^2/2T\}$. We can draw efficiently from this univariate distribution using rejection sampling with gaussian proposals.

Everything is now set up for applying the Exact Algorithm. We run the algorithm until we generate 5,000 exact skeletons of X . We had to propose 12,320 paths of $\mathbb{Z}$ to get the exact paths so we can estimate that the $P[I = 1]$ for the indicator of Theorem 2.1 is close to 0.41. In the 58% of the proposed paths the decision about accepting or rejecting a path was taken after simulating 1 and 2 points respectively uniformly from $(0, T) \times (0, T^{-1})$. The maximum number of points needed for the acceptance or the rejection of a path was 7 and 6 respectively.

Fig.2.3 shows on the left an exact skeleton of X and on the right the same skeleton after considering the transformation $x \mapsto \phi(x)$ for each of its joints. In the case of the ϕ -path the square encloses the area $(0, T) \times (0, T^{-1})$ and the black spots show the location of three points uniformly drawn from this rectangle. The numbers next to each circle show the order with which they were obtained. At

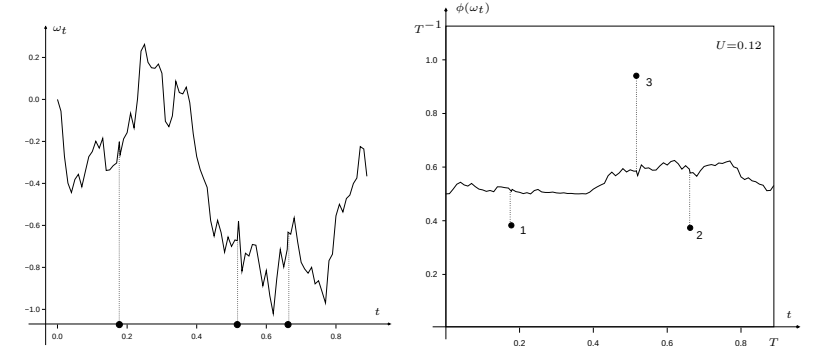


Figure 2.3: A case when the Exact Algorithm accepts the proposed path $\omega \sim \mathbb{Z}$.

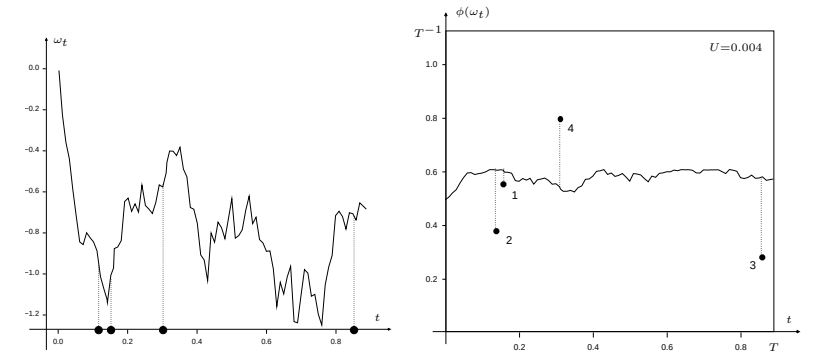


Figure 2.4: A case when the Exact Algorithm rejects a proposed path $\omega \sim \mathbb{Z}$.

the top right corner we have written the draw $U \sim \text{Un}(0, 1)$ needed by the Exact Algorithm. The third point exceeded the graph of ϕ so the algorithm decided that the event E_3 of Theorem 2.1 occurred after realising the proposed path ω only at three time instances. The rest of the path of X involves the time instances $0.01i$ for all $i = 1, 2, \dots$ such that $0.01i < T$ and is produced after filling in the proposed path.

Fig.2.4 shows similar graphs for the case when a proposed path is rejected. The graph that involves ϕ indicates that the proposed path was rejected because the

event E_4 did not occur.

Fig.2.5 shows the estimated density for the distribution of X_1 from samples of size 1,000,000 obtained after using the Exact Algorithm and the Euler approximation method for different time discretisations. For the case of the Exact Algorithm, to draw from X_1 we had to merge skeletons on the time intervals $[0,8/9]$ and $[8/9,1]$. Table 2.1 presents the times in seconds needed to get these samples and the p-values of the Kolmogorov-Smirnov test that compares the approximate draws of the Euler method with the exact ones of our algorithm. All programs were written in C and executed on a Athlon PC, running at 1500MHz. It is clear, at least for the case of the SDE given in (2.10), that the Exact Algorithm is superior to the Euler approximation even in terms of computational time.

Note also that, as implied in Proposition 2.6, the time needed to draw from X_l , $l > 0$, increases linearly in l for the Exact Algorithm. In contrast, the Euler method needs thinner approximation to produce reliable results as l increases because of the accumulating errors.

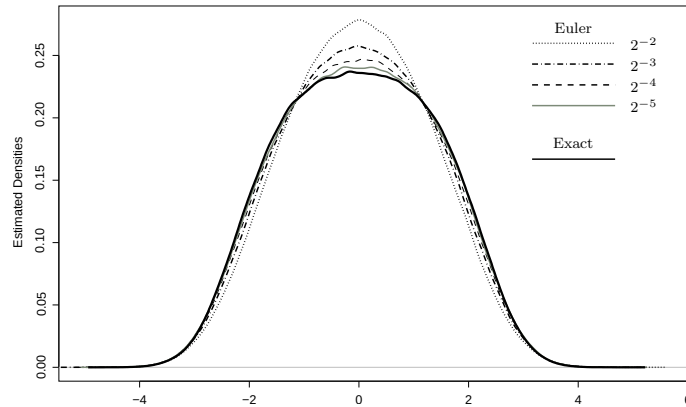


Figure 2.5: The estimated density for the distribution of X_1 from samples of size 1,000,000 generated by the Euler approximation (four cases for time increments 2^{-2} , 2^{-3} , 2^{-4} , 2^{-5}) and the Exact Algorithm.

Exact	Euler (for inc. $\Delta = 2^{-n}, n = 2, 3, \dots, 8$)	KS-test
35 sec	5 sec	0
	11 sec	0
	15 sec	0
	24 sec	0
	45 sec	0.002
	87 sec	0.261
	174 sec	0.412

Table 2.1: Times in seconds needed for the Exact Algorithm and the Euler approximation to produce 1,000,000 draws from the distribution of X_1 . The last column shows the p-values for the Kolmogorov-Smirnov test with null hypothesis that the exact and the corresponding approximate draws come from the same distribution.

2.6 Exact sampling of extremes and hitting times.

We take advantage of the properties of the Exact Algorithm to simulate exactly the maximum and the hitting time of a horizontal line boundary for one-dimensional diffusions determined by SDEs of the type (2.2) under the Conditions 1–3 given in Section 2.3.1.

As explained in Section 2.3.3, the Exact Algorithm reduces the problem of detecting characteristics for a path of the process X to the much more straightforward task of carrying out the certain detection process for its Brownian subpaths. To simplify the presentation of the algorithms that follow we denote by $\mathcal{S}_l\{x; t_1, t_2, \dots, t_n\}$ a skeleton of the target process X at the time instances $0 < t_1 < t_2 < \dots < t_n = l$ starting at x , for some length of interest $l > 0$.

For the case of the maximum value of X over the time interval $[0, l]$, we apply the Exact Algorithm, if necessary after dividing $[0, l]$ in smaller pieces of length at most the maximum length T permitted by the Exact Algorithm and dealing with each piece separately, and locate the maximum value of the accepted proposed paths after drawing the maximum of all the Brownian bridges (BBs) that intervene between the successive unveiled instances of these paths. Drawing the maximum of a BB is straightforward.

Let $W^y = \{W_s^y; s \in [0, t]\}$ be a Brownian motion over the interval $[0, t]$ started at y . If $M_t = \sup\{W_s^y; s \in [0, t]\}$ then it is a known result that:

$$P[M_t \in db | W_t^y = a] / db \propto (2b - y - a) \exp\left\{-\frac{(2b - y - a)^2}{2t}\right\}, \quad b \geq \max\{y, a\}.$$

For a proof, see for instance Karatzas and Shreve (1991), pg.95. This is just a linear transformation of a Rayleigh distribution and it is easy to verify that:

$$[M_t | W_t^y = a] \stackrel{d}{=} \frac{1}{2}(\sqrt{2tE + (a - y)^2} + a + y),$$

where $E \sim \text{Ex}(1)$. This formula generates the maximum of a BB of length t starting at y and finishing at a (for arbitrary t, y and a) and will be applied for all the BBs that fill in the exact skeleton of X . Thus, the complete algorithm for drawing the maximum of the target process $X = \{X_t; t \in [0, l]\}$ is as following:

Exact Simulation of the Maximum

1. Use the Exact Algorithm to draw a skeleton $\mathcal{S}_l\{0; t_1, t_2, \dots, t_n\}$.
 2. Simulate the maxima $M^1, M^2, \dots, M^n$ of the BBs that fill in $\mathcal{S}_l$.
 3. Output $\sup\{M^i; i = 1, 2, \dots, n\}$.
-

A similar procedure is carried out for the simulation of first passage times. We now have to check if each of the BBs that fill in the skeleton of X produced by the Exact Algorithm hit some arbitrary boundary γ or not and for the first bridge that hits γ to find the precise time instance when that occurs. Assume for simplicity that $\gamma > 0$ is greater than the starting point 0 of X .

The hitting time of a horizontal boundary for a BB is closely related with the first passage of an unconditional Brownian motion over a linear boundary. It is a known result that if $W(\delta) = \{W_s(\delta); s \in [0, t]\}$ is a BB of length t started at 0 and finishing at δ and $W = \{W_s; s \geq 0\}$ an unconditional Brownian motion

started at 0 then:

$$W_s(\delta) \stackrel{d}{=} \frac{s}{t}\delta + \frac{t-s}{\sqrt{t}}W_{s/(t-s)}, \quad s \in [0, t]. \quad (2.11)$$

Consider the boundary hitting times $\tau_\gamma(\delta) = \inf\{s \in [0, t] : W_s(\delta) \geq \gamma\}$ and $\tau_{\eta, \zeta} = \inf\{s \geq 0 : W_s \geq \eta + \zeta s\}$ with the convention that $\inf \emptyset = \infty$. Then for any $s \in [0, t]$ it is true that:

$$\begin{aligned} P[\tau_\gamma(\delta) > s] &= P[W_u(\delta) < \gamma, \text{ for all } 0 \leq u \leq s] = \\ &= P\left[\frac{u}{t}\delta + \frac{t-u}{\sqrt{t}}W_{u/(t-u)} < \gamma, \text{ for all } 0 \leq u \leq s\right] = \\ &= P\left[W_{u^*} < \frac{\gamma}{\sqrt{t}} + \frac{\gamma - \delta}{\sqrt{t}}u^*, \text{ for all } 0 \leq u^* \leq \frac{s}{t-s}\right] = \\ &= P\left[\tau_{\eta, \zeta} > \frac{s}{t-s}\right], \end{aligned}$$

for $\eta = \gamma/\sqrt{t}$, $\zeta = (\gamma - \delta)/\sqrt{t}$. It is now clear that for these values of η, ζ :

$$\tau_\gamma(\delta) \stackrel{d}{=} g(\tau_{\eta, \zeta}) \quad \text{for} \quad g(u) = \begin{cases} tu/(u+1), & 0 < u < \infty \\ \infty, & u = \infty \end{cases}. \quad (2.12)$$

The density of $\tau_{\eta, \zeta}$ is given by the Bachelier-Levy formula:

$$p^{\eta, \zeta}(u) = \frac{|\eta|}{u^{3/2}\sqrt{2\pi}} \exp\{-(\zeta u + \eta)^2/2u\}, \quad u > 0. \quad (2.13)$$

A proof can be found in Lerche (1986), ch.1. If we denote by $IGau(\mu, \lambda)$, $\mu > 0$, $\lambda > 0$, the inverse gaussian distribution with density:

$$IGau(\mu, \lambda, u) = \sqrt{\frac{\lambda}{2\pi u^3}} \exp\left\{-\frac{\lambda(u - \mu)^2}{2\mu^2 u}\right\}, \quad u > 0,$$

then, when $\eta\zeta < 0$ the density (2.13) is easily identified as an $IGau(-\eta/\zeta, \eta^2)$ and integrates to 1, i.e. the Brownian path hits the linear boundary $\eta + \zeta t$ with probability 1, as expected. When $\eta\zeta > 0$, the density (2.13) can be written as

$\exp(-2\zeta\eta) IGau(\eta/\zeta, \eta^2, u)$, which means that the Brownian path hits the linear boundary w.p. $\exp(-2\zeta\eta)$ and when it does the distribution of the hitting time is just the $IGau(\eta/\zeta, \eta^2)$. Draws from the inverse gaussian distribution can be generated in a very efficient way; see the algorithm described in Devroye (1986), ch.IV. Thus, simulating $\tau_{\eta, \zeta}$ and, via (2.12), $\tau_\gamma(\delta)$ for any values of the involved parameters is straightforward.

We now have all that is needed to carry out the algorithm for the exact simulation of the hitting time of a horizontal boundary for the solution X of (2.2). Since it is not possible to know a-priori the length of the path of X we need to construct before we locate the time instance when X hits the boundary we merge as many exact skeletons of X of eligible length T as necessary before some intervening BB hits the boundary.

Exact Simulation of Hitting Time

1. Set $i = 1$, $y = 0$.
 2. Use the Exact Algorithm to generate $\mathcal{S}_T\{y; t_{i,1}, t_{i,2}, \dots, t_{i,n_i}\}$.
 - 3₁ If the intervening BBs hit γ , save the first time τ when that occurs.
 - 3₂ Else, set $y = X_T$, $i = i + 1$ and goto (2).
 4. Return $(i - 1)T + \tau$.
-

Assume that N^γ is the total number of the uniformly drawn points needed for finding $\tau_\gamma = \inf\{t \geq 0 : X_t = \gamma\}$ and that N_i is the number of points needed for accepting a proposed path from $(i-1)T$ to iT , $i \geq 1$. Let $\mathcal{G}_i$ be the information for the obtained exact path of X until the time iT , $i \geq 0$. Proposition 2.6 is true for any given starting point of the target process, so $\mathbb{E}[N_i | \mathcal{G}_{i-1}] \leq e^2$ for all $i \geq 1$. We set $\tau := \lceil \tau_\gamma / T \rceil$; τ is the total number of the exact skeletons we have to merge

before we find τ_γ . Clearly,

$$\begin{aligned} \mathbb{E}[N^\gamma] &= \sum_{i=1}^{\infty} \mathbb{E}[N_i \cdot \mathbb{I}\{\tau \geq i\}] = \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{E}[N_i \cdot \mathbb{I}\{\tau \geq i\} | \mathcal{G}_{i-1}]] = \\ &= \sum_{i=1}^{\infty} \mathbb{E}[\mathbb{I}\{\tau \geq i\} \cdot \mathbb{E}[N_i | \mathcal{G}_{i-1}]] \leq e^2 \sum_{i=1}^{\infty} \mathbb{P}[\tau \geq i] \leq \\ &\leq e^2 \left(\frac{\mathbb{E}[\tau_\gamma]}{T} + 1 \right), \end{aligned}$$

where we have used the monotone convergence theorem and the fact that $\{\tau \geq i\}$ is $\mathcal{G}_{i-1}$ -measurable. So the expected time for the termination of the algorithm, in terms of the uniformly drawn points needed, is finite when $\mathbb{E}[\tau_\gamma]$ is finite.

2.6.1 An application

We applied the above exact algorithms to the solution X of the SDE (2.10) considered in Section 2.5. We compare the exact draws of our algorithms with the approximate ones of the simple Euler scheme that considers the continuous time process $Y = \{Y_t\}_{t \geq 0}$, $Y_0 = 0$, defined on the instances $\{ih\}_{i \geq 1}$, for some chosen increment $h > 0$, via the recursion

$$Y_{(i+1)h} = Y_{ih} + \sin(Y_{ih})h + \sqrt{h}Z_{i+1},$$

where Z_{i+1} , $i \geq 0$, are iid with $Z_1 \sim \mathcal{N}(0, 1)$. The paths of Y become continuous after considering the linear interpolations between the successive instances of the grid $\{ih\}_{i \geq 1}$.

On the left of Fig.2.6, we show a qq-plot comparing two samples each of size 50,000 from the distribution of $M_2^X = \sup\{X_t; t \in [0, 2]\}$ generated by the Exact Algorithm and the Euler approximation. For the Euler scheme we used increments $h = 2^{-12}$. On the right of the same figure, we show the times in seconds needed to get 50,000 draws from M_2^X for the Exact Algorithm and the Euler scheme for

different increments h . We have also included the results of a Kolmogorov-Smirnov test that compares the approximate samples of the Euler scheme with the exact sample as an indication of how small h has to be for the Euler scheme to give correct results.

In Fig.2.7 we present similar results for the case of the simulation of the first passage time $\tau_2 = \inf\{t \geq 0 : X_t = 2\}$. Actually we tune the algorithm to obtain draws from the $\min\{\tau_2, 10\}$ since it can be shown that $\mathbb{E}[\tau_2] = \infty$ so the expected number of the uniformly drawn points for the completion of the algorithm for drawing from τ_2 is infinite.

It is remarkable that in both cases the Exact Algorithm appears much more efficient than the Euler approximation.

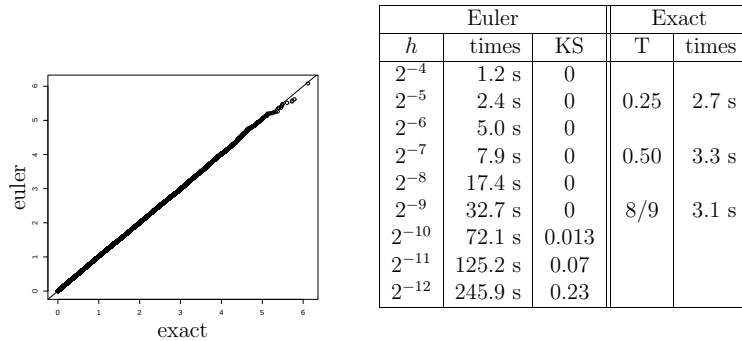


Figure 2.6: Results from the simulation of the maximum $M_2^X = \sup\{X_t; t \in [0, 2]\}$ of the solution of (2.10).

2.7 Extensions and conclusions

In this paper we have introduced a simple but computationally effective way of simulating exactly from a family of diffusion processes. The algorithm outputs a skeleton which can be readily “filled in” as and when necessary using simple gaussian random variables, and crucially independently of the diffusion we are

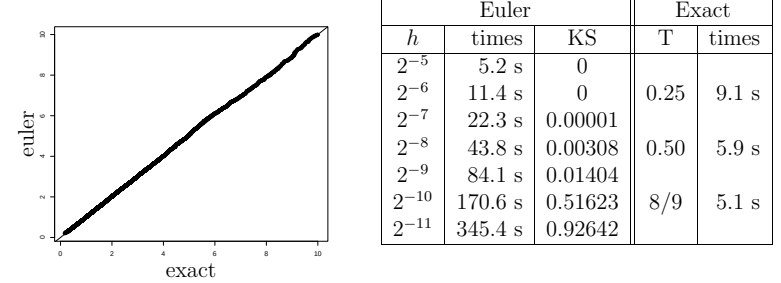


Figure 2.7: Similar results with those of Fig.2.6 for the case of the simulation of $\min\{\tau_2, 10\}$ for $\tau_2 = \inf\{t \geq 0 : X_t = 2\}$.

attempting to simulate from.

We have not carried out an extensive simulations study. However in the examples considered, our method performs very favourably in comparison to the obvious numerical approximation alternative using the Euler scheme. In Section 2.4 we also give results which show that the method can be robust to the length of the time-series, and the computing time is at worst linear in the degree of non-linearity (as measured by the range of $\alpha^2 + \alpha'$).

The convenient form of the output allows the algorithm to be used in a number of ways including the construction of reduced variance Monte Carlo estimation. In Section 2.5 we discuss unbiased estimation of boundary hitting times and diffusion maxima. We envisage future application in inference for stochastic processes and finance.

The most demanding of the Conditions 1–3 (detailed in Section 2.3.1) required for the algorithm to work, is that the functional $\alpha^2 + \alpha'$ of the drift be bounded. In most cases $\alpha^2 + \alpha'$ is bounded from below but not from above so we can still produce (2.5) for some $\phi \geq 0$ and hope for a valid rejection sampling scheme. Ongoing work is investigating such an approach.

It is worth remarking that the approach outlined in this paper extends routinely to SDEs for non-Markov processes absolutely continuous w.r.t. appropriate mar-

tingales, and our focus on diffusions has been purely for simplicity. Furthermore, it is easy to extend these results some way towards considering jump diffusions.

Acknowledgements

We thank Professor Duncan Murdoch for stimulating discussions on the problem of the exact simulation of diffusions. We also thank the referees for their useful suggestions. One of the authors research was funded by the Greek State Scholarship's Foundation.

APPENDIX

Proof of Proposition 2.1 (Rejection Sampling):

Note that for any $i = 1, 2, \dots$ we get:

$$\mathbb{P}[I_i = 1] = \int_C \mathbb{P}[I_i = 1 | Y_i = y] \nu(dy) = \int_C f(y) \nu(dy) = \int_C \epsilon \mu(dy) = \epsilon .$$

Trivially, for any $F \in \mathcal{C}$ we have the property:

$$\mathbb{P}[Y_\tau \in F] = \mathbb{P}[Y_\tau \in F, I_1 = 1] + \mathbb{P}[Y_\tau \in F | I_1 = 0] \cdot \mathbb{P}[I_1 = 0] . \quad (2.14)$$

From the independence among the members of the sequence $(Y_n, I_n)_{n \geq 1}$ it is clear that:

$$\mathbb{P}[Y_\tau \in F | I_1 = 0] = \mathbb{P}[Y_\tau \in F] . \quad (2.15)$$

We can easily find the $\mathbb{P}[Y_\tau \in F, I_1 = 1]$ in the following way:

$$\mathbb{P}[Y_\tau \in F, I_1 = 1] = \int_F \mathbb{P}[I_1 = 1 | Y_1 = y] \nu(dy) = \int_F f(y) \nu(dy) = \epsilon \mu(F) . \quad (2.16)$$

From (2.14) using (2.15) and (2.16) we get:

$$\mathbb{P}[Y_\tau \in F] = \epsilon \mu(F) + (1 - \epsilon) \mathbb{P}[Y_\tau \in F] ,$$

which yields that $\mathbb{P}[Y_\tau \in F] = \mu(F)$.

Proof of Proposition 2.3 (Biased Brownian Motion):

Choose any $F \in \mathcal{C}$. We will show that $\mathbb{E}_{\mathbb{Z}}[\mathbb{I}_F] = \mathbb{E}_{\mathbb{W}}[\mathbb{I}_F f]$ where we have set:

$$f(\omega) := \frac{h(\omega_T) \sqrt{2\pi T}}{\exp\left(-\frac{\omega_T^2}{2T}\right)}, \quad \omega \in C .$$

Let $\mathcal{C}_T$ be the σ -algebra corresponding to the information about the ending point of the paths. Note that f is $\mathcal{C}_T$ -measurable. From the definition of $\mathbb{Z}$ it is clear that:

$$\mathbb{Z}[F | \mathcal{C}_T] = \mathbb{W}[F | \mathcal{C}_T] =: g(\omega_T), \quad \mathbb{W}\text{-a.s.} ,$$

for some measurable function $g : \mathbf{R} \mapsto \mathbf{R}$. Clearly,

$$\begin{aligned} \mathbb{E}_{\mathbb{W}}[\mathbb{I}_F f] &= \mathbb{E}_{\mathbb{W}}[\mathbb{E}_{\mathbb{W}}[\mathbb{I}_F f | \mathcal{C}_T]] = \mathbb{E}_{\mathbb{W}}[f \mathbb{W}[F | \mathcal{C}_T]] = \\ &= \int_{\mathbf{R}} \frac{h(u) \sqrt{2\pi T} \exp\left(-\frac{u^2}{2T}\right)}{\sqrt{2\pi T} \exp\left(-\frac{u^2}{2T}\right)} g(u) du = \int_{\mathbf{R}} h(u) g(u) du , \end{aligned}$$

since w.r.t. $\mathbb{W}$, $\omega_T \sim \mathcal{N}(0, T)$. It is straightforward that:

$$\mathbb{E}_{\mathbb{Z}}[\mathbb{I}_F] = \mathbb{E}_{\mathbb{Z}}[\mathbb{E}_{\mathbb{Z}}[\mathbb{I}_F | \mathcal{C}_T]] = \int_{\mathbf{R}} h(u) g(u) du ,$$

since under $\mathbb{Z}$ we know that $\omega_T \sim h$. It is clear that $\mathbb{E}_{\mathbb{Z}}[\mathbb{I}_F] = \mathbb{E}_{\mathbb{W}}[\mathbb{I}_F f]$.

Proof of Theorem 2.1:

(i) From (2.8) it is straightforward that $E_1 \subseteq E_3 \subseteq \dots \subseteq E_{2n+1} \subseteq E_{2n+2}$ for

any $n \in \{0, 1, \dots\}$. From the definition of $(E_{2n+2})_{n \geq 0}$ we can conclude that $E_{2n} = E_{2n+2} \cup (\Gamma_{2n+1} - \Gamma_{2n+2})$, so $E_{2n+2} \subseteq E_{2n} \subseteq \dots \subseteq E_2$. Clearly, $E_{2\kappa+1} \subseteq E_{2\lambda+2}$ for any $\kappa, \lambda \in \{0, 1, \dots\}$.

The definition of $(\Gamma_n)_{n \geq 0}$ implies that for any $n \geq 0$ it is true that $\Gamma_{n+1} \subseteq \Gamma_n$, $\bigcap_n \Gamma_n$ is a set of zero probability and $\Gamma_{2n+2} = E_{2n+2} - E_{2n+1}$. Therefore:

$$\left(\bigcap_n E_{2n+2}\right) - \left(\bigcup_n E_{2n+1}\right) = \bigcap_n (E_{2n+2} - E_{2n+1}) = \bigcap_n \Gamma_{2n+2}.$$

Trivially, $\bigcup_n E_{2n+1} \subseteq \bigcap_n E_{2n+2}$ and their difference $\bigcap_n E_{2n+2} - \bigcup_n E_{2n+1}$ has zero probability.

(ii) Since $E_{2n+1} \uparrow E$ it is true that:

$$\mathbb{P}[I = 1 \mid \omega] \equiv \mathbb{P}[E \mid \omega] = \lim_{n \rightarrow \infty} \mathbb{P}[E_{2n+1} \mid \omega]. \quad (2.17)$$

Recall that $E_{2n+1} = \sum_{k=0}^n (\Gamma_{2k} - \Gamma_{2k+1})$. We can now get that:

$$\mathbb{P}[E_{2n+1} \mid \omega] = \sum_{k=0}^n (\mathbb{P}[\Gamma_{2k} \mid \omega] - \mathbb{P}[\Gamma_{2k+1} \mid \omega]) = \sum_{k=0}^{2n+1} (-1)^k \mathbb{P}[\Gamma_k \mid \omega]. \quad (2.18)$$

From the fact that $\tau = (V_n, W_n)_{n \geq 1}$ are iid and τ , U , and ω are independent we conclude that:

$$\mathbb{P}[\Gamma_k \mid \omega] = \mathbb{P}\left[U \leq \frac{1}{k!} \mid \omega\right] \prod_{i=1}^k \mathbb{P}[\phi(\omega_{V_i}) \geq W_i \mid \omega]. \quad (2.19)$$

The $\mathbb{P}[\phi(\omega_{V_i}) \geq W_i \mid \omega]$ is the probability that given a path $\omega \sim \mathbb{Z}$ a point uniformly selected from the rectangle $(0, T) \times (0, 1/T)$ is found below the graph $\{(t, \phi(\omega_t)); t \in [0, T]\}$. From (2.6) and elementary probability theory we get:

$$\mathbb{P}[\phi(\omega_{V_k}) \geq W_k \mid \omega] = \int_0^T \phi(\omega_t) dt.$$

It is clear that $\mathbb{P}[U \leq \frac{1}{k!} \mid \omega] = \frac{1}{k!}$, so (2.19) yields:

$$\mathbb{P}[\Gamma_k \mid \omega] = \frac{1}{k!} \left\{ \int_0^T \phi(\omega_t) dt \right\}^k. \quad (2.20)$$

Starting from (2.17) and using (2.18), (2.20) we get:

$$\mathbb{P}[I = 1 \mid \omega] = \lim_{n \rightarrow \infty} \sum_{k=0}^{2n+1} \frac{(-1)^k}{k!} \left\{ \int_0^T \phi(\omega_t) dt \right\}^k = \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\}.$$

Proof of Proposition 2.4:

Since $0 \leq \phi \leq T^{-1}$, the probability of accepting an arbitrary path $\omega \sim \mathbb{Z}$ will be $\exp\{-\int_0^T \phi(\omega_t) dt\} \geq e^{-1}$.

Recall that T is restricted not to be bigger than some constant $T_0 (= \frac{1}{k_2 - k_1})$, see Condition 3 for the definition of k_1, k_2 . To emphasise the involvement of the time variable in what follows we define $p : [0, T_0] \times C \mapsto [0, 1]$ with:

$$p(T, \omega) = \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\}.$$

From result (ii) of Theorem 2.1 it is clear that:

$$\epsilon(T) \equiv \int_C p(T, \omega) \mathbb{Z}(d\omega) \quad (2.21)$$

for all eligible T . Recall that $0 \leq \phi \leq T^{-1}$ for any $T \leq T_0$. Trivially, p is decreasing in T and $\lim_{T \downarrow 0} p(T, \omega) = 1$, both properties being true $\mathbb{Z}$ -a.s.. The first property yields, after using (2.21), that $\epsilon(T)$ is decreasing in T . Also, since $\lim_{n \rightarrow \infty} p(1/n, \omega) = 1$ and $p(\frac{1}{n}, \omega) \leq p(\frac{1}{n+1}, \omega)$ for any $n \geq 1$ the monotone convergence theorem gives:

$$\lim_{n \rightarrow \infty} \epsilon(1/n) = \lim_{n \rightarrow \infty} \int_C p(1/n, \omega) \mathbb{Z}(d\omega) = \int_C \left\{ \lim_{n \rightarrow \infty} p(1/n, \omega) \right\} \mathbb{Z}(d\omega) = 1.$$

From the monotonicity of $\epsilon(T)$ we conclude that $\lim_{T \downarrow 0} \epsilon(T) = 1$.

Proof of Proposition 2.6:

For the time interval $[0, T]$, where $T = 1/R$, the Exact Algorithm uses in total $N_T := N_1 + N_2 + \dots + N_J$ points, where N_i is the number of points needed to decide about the i -th proposed path and J is the number of the proposed paths until one is accepted, $J = 1, 2, \dots$ and $i = 1, 2, \dots, J$. We denote by N the number of points for deciding (in general) about a path $\omega \sim \mathbb{Z}$ and by $\mathbb{E}[N | A]$ and $\mathbb{E}[N | A^c]$ the expected number of points for deciding about a path $\omega \sim \mathbb{Z}$ given that the path is accepted and rejected respectively. Let $\epsilon = \mathbb{P}[A]$ be the probability of accepting a path. Then:

$$\begin{aligned} \mathbb{E}[N_T] &= \mathbb{E}[\mathbb{E}[N_T | J]] = \mathbb{E}\left[\sum_{i=1}^{J-1} \mathbb{E}[N_i | J] + \mathbb{E}[N_J | J]\right] = \\ &= \mathbb{E}[(J-1) \cdot \mathbb{E}[N | A^c] + \mathbb{E}[N | A]] = \\ &= (1/\epsilon - 1) \mathbb{E}[N | A^c] + \mathbb{E}[N | A] = \frac{1}{\epsilon} \cdot \mathbb{E}[N]. \end{aligned}$$

From Proposition 2.4 we have that $1/\epsilon \leq e$ and from Proposition 2.5, $\mathbb{E}[N] \leq e$, so $\mathbb{E}[N_T] \leq e^2$. The same result can be obtained for the any $\lceil l \cdot R \rceil$ skeletons of length $T = 1/R$ (or less, for the case of the last skeleton when $1/R$ does not divide l) that will merge to construct the complete skeleton of length l .

Article 3

Retrospective Exact Simulation of Diffusion Sample Paths with Applications

3.1 Introduction

Applications of diffusion models are ubiquitous throughout science; a representative list might include finance, biology, physics and engineering. In many cases the evolution of some phenomenon is described by a scalar stochastic process $X = \{X_t; t \in [0, T]\}$ determined as the solution of an SDE of the type:

$$dX_t = b(X_t) dt + \sigma(X_t) dW_t, \quad t \in [0, T], \quad X_0 = x \in \mathbf{R}, \quad (3.1)$$

driven by the Brownian motion $\{W_t; t \in [0, T]\}$. The drift b and the coefficient function σ are presumed to satisfy the regularity conditions (locally Lipschitz, with a growth bound) that guarantee the existence of a weakly unique, global solution of (3.1), see ch.4 of Kloeden and Platen (1992). In fact, we will restrict our attention

to SDEs of the type:

$$dX_t = \alpha(X_t) dt + dW_t, \quad t \in [0, T], \quad X_0 = x \in \mathbf{R}, \quad (3.2)$$

for some drift function α , since (3.1) can be transformed into an SDE of unit coefficient under mild additional conditions on σ , by applying the transformation $X_t \mapsto \eta(X_t)$, where

$$\eta(x) = \int_z^x \frac{1}{\sigma(u)} du, \quad (3.3)$$

with z some element of the state space of X .

Simulation and inference for SDEs of the form (3.1) generally require some kind of discrete approximation (see for example Roberts and Stramer, 2001; Kloeden and Platen, 1992). To this cause, an Euler approximation of the SDE might be used:

$$X_{t+\Delta} = X_t + \alpha(X_t)\Delta + \mathcal{N}(0, \Delta).$$

Such methods are correct only in an asymptotic sense (i.e. as the time discretisation gets finer). When applicable, our exact simulation algorithm, termed the Exact Algorithm, returns skeletons of *exact* paths of X .

The Exact Algorithm carries out rejection sampling using Brownian paths as proposals, and returns skeletons of the target SDE (3.2) obtained at some random time instances. The skeletons can be later filled in *independently* of X by interpolation of Brownian bridges. An initial version of the Exact Algorithm was given in Beskos and Roberts (2005) [Article 2]. Its most demanding restriction was that the functional $\alpha^2 + \alpha'$ of the drift be bounded. In this paper we describe an easier and more flexible method to apply the rejection sampling algorithm proposed in Beskos and Roberts (2005) [Article 2], but most importantly we substantially relax the boundedness condition on the drift. Assuming that $\alpha^2 + \alpha'$ is bounded from below, we now require that $\limsup_{u \rightarrow \infty} (\alpha^2 + \alpha')(u)$ and $\limsup_{u \rightarrow -\infty} (\alpha^2 + \alpha')(u)$ are not both $+\infty$. Thus, the method introduced in this paper can now be applied

to a substantially wider class of diffusion processes.

A major appeal of the Exact Algorithm is that it can be adapted to provide solutions to a variety of challenging problems related to diffusion processes, and in this paper we explore some of these possibilities. Maximum likelihood inference for discretely observed diffusions is known to be particularly difficult, even when the drift and diffusion coefficients are restricted to small-dimensional parametric classes. This problem has received considerable attention in the last two decades from the statistical, applied probability and econometric community, see for example Sørensen (2004) for a recent review and references. All current methods are subject to approximation error, and are often either difficult to implement or subject to substantial Monte Carlo error. We show how our simulation algorithm can be easily used to provide efficient Monte Carlo MLEs (maximum likelihood estimators).

The Exact Algorithm can also be used to simulate diffusion processes conditioned to hit specific values at predetermined time instances (diffusion bridges). Contrary to all other current simulation methods, with our approach the conditional simulation is in fact easier than the unconditional simulation.

The structure of the paper is as follows. In Section 3.2 we present the Exact Algorithm. In Section 3.3 we describe analytically the extension noted above; it requires the stochastic calculus theory that makes possible a simulation-oriented decomposition of a Brownian path at its minimum. Section 3.4 presents some results related with the efficiency of the algorithm. In Section 3.5 we apply our algorithm to simulate exactly from the otherwise intractable logistic growth model. In Section 3.6 we demonstrate the use of the method for the simulation of conditional diffusions and parametric inference. We finish in Section 3.7 with some general remarks about the Exact Algorithm and possible extensions in future research.

3.2 Retrospective rejection sampling

Before we introduce the basic form of the Exact Algorithm we require the following preliminary notation. Let $C \equiv C([0, T], \mathbf{R})$ be the set of continuous mappings from $[0, T]$ to $\mathbf{R}$ and ω be a typical element of C . We equip C with the cylinder σ -algebra $\mathcal{C}$ generated by the co-ordinate mappings. We denote by $W^x = \{W_t^x; t \in [0, T]\}$ the Brownian motion started at $x \in \mathbf{R}$.

Let $\mathbb{Q}$ be the probability measure on $(C, \mathcal{C})$ corresponding to the law of the solution X of (3.2) and $\mathbb{W}$ the corresponding probability measure for W^x . The objective is to construct a rejection sampling algorithm to draw from $\mathbb{Q}$. The Girsanov transformation of measures (see for instance Oksendal (1998), ch.8) implies that:

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(\omega) = \exp \left\{ \int_0^T \alpha(\omega_t) d\omega_t - \frac{1}{2} \int_0^T \alpha^2(\omega_t) dt \right\}.$$

Under the condition that α is everywhere continuously differentiable we can eliminate the Itô's integral after applying Itô's lemma to $A(\omega_t)$ for $A(u) := \int_0^u \alpha(y) dy$, $u \in \mathbf{R}$. Simple calculations give:

$$\frac{d\mathbb{Q}}{d\mathbb{W}}(\omega) = \exp \left\{ A(\omega_T) - A(x) - \frac{1}{2} \int_0^T \left(\alpha^2(\omega_t) + \alpha'(\omega_t) \right) dt \right\}.$$

To remove the possible inconvenience of A being unbounded and in any case to simplify the Radon-Nikodym derivative of the two probability measures, we will use candidate paths from a process identical to W^x except for the distribution of its ending point. We call such a process a *biased* Brownian motion.

We thus consider the biased Brownian motion defined as the conditioned Wiener process $(W^x | W_T^x \sim h)$ for a density h proportional to $\exp\{A(u) - (u - x)^2/2T\}$, $u \in \mathbf{R}$. It is now necessary that h is integrable. Let $\mathbb{Z}$ be the probability measure on $(C, \mathcal{C})$ corresponding to the law of the biased Brownian motion defined above. Note that it is straightforward to simulate a skeleton of $\omega \sim \mathbb{Z}$ just by simulating first its ending point $\omega_T \sim h$. The measures $\mathbb{Z}$ and $\mathbb{W}$ are equivalent and their

Radon-Nikodym derivative can be easily derived from the following proposition:

Proposition 3.1 *Let $M = \{M_t; t \in [0, T]\}$, $N = \{N_t; t \in [0, T]\}$ be two stochastic processes on $(C, \mathcal{C})$ with corresponding probability measures $\mathbb{M}, \mathbb{N}$. Assume that f_M, f_N are the densities of the ending points M_T and N_T respectively with identical support $\mathbf{R}$. If it is true that $(M | M_T = \rho) \stackrel{d}{=} (N | N_T = \rho)$, for all $\rho \in \mathbf{R}$, then:*

$$\frac{d\mathbb{M}}{d\mathbb{N}}(\omega) = \frac{f_M(\omega_T)}{f_N(\omega_T)}.$$

Proof: In the Appendix.

Therefore, $d\mathbb{W}/d\mathbb{Z}(\omega) = \mathcal{N}_{x,T}/h(\omega_T) \propto \exp\{-A(\omega_T)\}$, where $\mathcal{N}_{x,T}$ represents the density of the normal distribution with mean x and variance T , and:

$$\frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) = \frac{d\mathbb{Q}}{d\mathbb{W}}(\omega) \frac{d\mathbb{W}}{d\mathbb{Z}}(\omega) \propto \exp \left\{ - \int_0^T \left(\frac{1}{2} \alpha^2(\omega_t) + \frac{1}{2} \alpha'(\omega_t) \right) dt \right\}.$$

Assume now that $(\alpha^2 + \alpha')/2$ is bounded below. Then, we can obtain a non-negative function ϕ such that:

$$\frac{d\mathbb{Q}}{d\mathbb{Z}}(\omega) \propto \exp \left\{ - \int_0^T \phi(\omega_t) dt \right\} \leq 1, \quad \mathbb{Z} - \text{a.s.} \quad (3.4)$$

Analytically, ϕ is defined as:

$$\phi(u) = \frac{\alpha^2(u) + \alpha'(u)}{2} - k, \quad u \in \mathbf{R}, \quad \text{for a fixed } k \leq \inf_{u \in \mathbf{R}} (\alpha^2 + \alpha')(u)/2. \quad (3.5)$$

We now summarise the three conditions that allow the derivation of (3.4).

1. The drift function α is continuously differentiable.
2. The function $\exp\{A(u) - (u - x)^2/2T\}$, $u \in \mathbf{R}$, for $A(u) = \int_0^u \alpha(y) dy$, is integrable.

3. The function $(\alpha^2 + \alpha')/2$ is bounded below.

In a simulation context, were it possible to draw complete continuous paths $\omega \sim \mathbb{Z}$ on $[0, T]$ and calculate the integral involved in (3.4) analytically then rejection sampling on the probability measures $\mathbb{Q}, \mathbb{Z}$ would be straightforward. The Exact Algorithm manages to circumvent these difficulties in order to carry out exact rejection sampling using only finite information about the proposed paths from $\mathbb{Z}$. Given a proposed path $\omega \sim \mathbb{Z}$, it defines an event of probability $\exp\{-\int_0^T \phi(\omega_t)dt\}$. The truth or falsity of this event can be determined after unveiling ω only at a finite collection of time instances. The rejection sampling scheme is carried out in a *retrospective* way since the realisation of the proposed variate (the path $\omega \sim \mathbb{Z}$) at some required instances follows that of the variates that decide for the acceptance or not of this proposed variate. The idea of retrospective sampling, in a different context, has been introduced in Papaspiliopoulos and Roberts (2004) where it is applied to an MCMC algorithm for Bayesian analysis from a Dirichlet mixture model.

We will now present a simple theorem that demonstrates the idea behind the method.

Theorem 3.1 *Let ω be any element of $C([0, T], \mathbf{R})$ and $r(\omega)$ an upper bound for the mapping $t \mapsto \phi(\omega_t)$, $t \in [0, T]$. If Φ is a homogeneous Poisson process of unit intensity on $[0, T] \times [0, r(\omega)]$ and $N =$ number of points of Φ found below the graph $\{(t, \phi(\omega_t)); t \in [0, T]\}$, then:*

$$P[N = 0 | \omega] = \exp\left\{-\int_0^T \phi(\omega_t)dt\right\}.$$

Proof: Conditionally on ω , N follows a Poisson distribution with mean $\int_0^T \phi(\omega_t)dt$.

□

Theorem 3.1 suggests that were it possible to generate complete, continuous paths $\omega \sim \mathbb{Z}$ then we could carry out rejection sampling on $\mathbb{Q}, \mathbb{Z}$ without having to

evaluate the integral involved in (3.4); we would only have to generate a realisation of the Poisson process Φ and check if all the points of Φ fall above the ϕ -graph to decide about the acceptance (the rejection in any other case) of ω .

The simple observation that only finite information about some $\omega \sim \mathbb{Z}$ suffices for the algorithm to decide about it will yield a simple, valid rejection sampling scheme which involves finite computations. The technical difficulty of locating a rectangle for the realisation of the Poisson process, or equivalently an upper bound for the map $t \mapsto \phi(\omega_t)$, imposes some restrictions on the applicability of the algorithm.

3.2.1 The case when ϕ is bounded

Let r be an upper bound of ϕ . This is the simple case and is similar to the one considered in Beskos and Roberts (2005) [Article 2]. We now devise a different approach which is simpler and even more efficient than the one presented in Beskos and Roberts (2005) [Article 2]. Moreover this method is particularly useful for the more general case of Section 3.3. Based on Theorem 3.1 and (3.4) we can generate a feasible rejection sampling scheme just by thinking retrospectively, i.e. first realise the Poisson process and then construct the path $\omega \sim \mathbb{Z}$ only at the time instances required to determine N .

Exact Algorithm 1 (EA1)

1. Produce a realisation $\{(\psi_1, v_1), (\psi_2, v_2), \dots, (\psi_\kappa, v_\kappa)\}$, of the Poisson process Φ on $[0, T] \times [0, r]$,
 2. Simulate a skeleton of $\omega \sim \mathbb{Z}$ at the times $\{\psi_1, \psi_2, \dots, \psi_\kappa\}$.
 3. Evaluate N .
 4. If $N = 0$ goto (5), else goto (1).
 5. Output the currently constructed skeleton $S(\omega)$ of ω .
-

This algorithm returns an exact skeleton of X and is shown (see Section 3.4) to run in finite time. In Section 3.4 we present details about the efficiency of the algorithm. When a skeleton $S(\omega)$ is accepted as a realisation from $\mathbb{Q}$, we can continue constructing it according to $\mathbb{Z}$, i.e. using Brownian bridges between the successive currently unveiled instances of ω . In this way, the corresponding path of $\mathbb{Q}$ can be realised at any requested time instances.

Fig.3.1 illustrates the retrospective idea behind EA1. The graph on the right emphasises the fact that skeletons can be readily filled in using independent Brownian bridges and without further reference to the dynamics of the target process X . This is precisely the property we shall exploit in Section 3.6.1 to carry out maximum likelihood parametric inference.

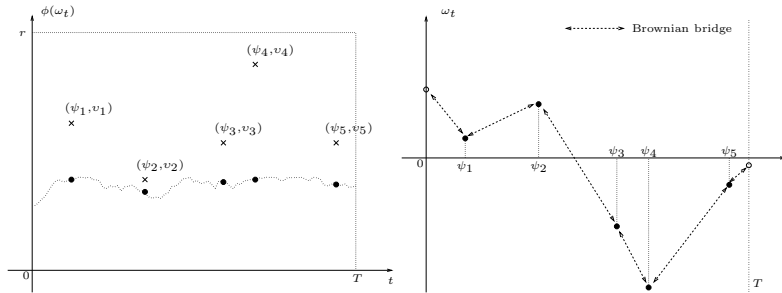


Figure 3.1: The retrospective idea: the signs ($\times$) show the position of the points from a realisation of Φ . The black dots on the left graph show the only instances of the path $\phi(\omega_t)$ we needed to realise before accepting (since $N = 0$) the corresponding unveiled instances of ω (the black dots on the right graph) as an exact skeleton of X . Intermediate points of the skeleton are obtained via Brownian bridges between successive instances of it.

3.3 The case when $\limsup_{u \rightarrow \infty} \phi(u) < \infty$

It is well-known that it is possible to construct a Brownian path (or a biased Brownian path) on a bounded time interval after simulating first its minimum (or

maximum) and then the rest of the path using Bessel processes. For a definition and properties of the Bessel process see for instance Karatzas and Shreve (1991) and Revuz and Yor (1991).

We will exploit this decomposition of the Brownian path to present an extension of EA1 when either $\limsup_{u \rightarrow \infty} \phi(u) < \infty$ or $\limsup_{u \rightarrow -\infty} \phi(u) < \infty$. Without loss of generality, we will consider the case when $\limsup_{u \rightarrow \infty} \phi(u) < \infty$; it is then possible to identify an upper bound $r(\omega)$ for the mapping $t \mapsto \phi(\omega_t)$, $t \in [0, T]$, after decomposing the proposed path ω at its minimum. For an arbitrary $\omega \in C$ we define

$$m = \inf\{\omega_t; t \in [0, T]\}.$$

We can now set

$$r(\omega) \equiv r(m) = \sup\{\phi(u); u \geq m\} < \infty. \quad (3.6)$$

By symmetry, the same theory covers the case when $\limsup_{u \rightarrow -\infty} \phi(u) < \infty$. To describe the decomposition of a Brownian path at its minimum we need to recall some properties of the Brownian motion.

3.3.1 Decomposing the Brownian path at its minimum

We will need the time of the minimum for a given path:

$$\tau = \sup\{t \in [0, T] : \omega_t = m\}.$$

Let $\omega \sim \mathbb{W}$ be a Brownian motion path on $[0, T]$ started at 0. Then, for any fixed point $a \in \mathbf{R}$:

$$\mathbf{P}[m \in db, \tau \in dt | \omega_T = a] \propto \frac{b(b-a)}{\sqrt{t^3(T-t)^3}} \exp\left\{-\frac{b^2}{2t} - \frac{(b-a)^2}{2(T-t)}\right\} db dt, \quad (3.7)$$

with $b \leq \min\{a, 0\}$ and $t \in [0, T]$. For the derivation of this distribution see for instance ch.2 of Karatzas and Shreve (1991). In the following proposition we

describe a simple algorithm for drawing from (3.7). We denote by $IGau(\mu, \lambda)$, $\mu > 0$, $\lambda > 0$, the inverse gaussian distribution with density:

$$IGau(\mu, \lambda, u) = \sqrt{\frac{\lambda}{2\pi u^3}} \exp \left\{ -\frac{\lambda(u - \mu)^2}{2\mu^2 u} \right\}, \quad u > 0.$$

A simple algorithm for drawing from this density is described in ch.IV of Devroye (1986).

Proposition 3.2 *Let $E \sim \text{Ex}(1)$ and $Z_1 = (a - \sqrt{2TE + a^2})/2$. If $Z_1 = b$ is a realisation of Z_1 , set $c_1 = (a - b)^2/2T$, $c_2 = b^2/2T$. Let $U \sim \text{Un}(0, 1)$, $I_1 \sim IGau(\sqrt{c_1/c_2}, 2c_1)$ and $I_2 \sim 1/IGau(\sqrt{c_2/c_1}, 2c_2)$ independently and define:*

$$V = \mathbb{I} [U < (1 + \sqrt{c_1/c_2})^{-1}] \cdot I_1 + \mathbb{I} [U \geq (1 + \sqrt{c_1/c_2})^{-1}] \cdot I_2.$$

Then the pair (Z_1, Z_2) , for $Z_2 := T/(1 + V)$, is distributed according to (3.7).

Proof: In the Appendix.

Recall from the definition of $\mathbb{Z}$ that to construct a path $\omega \sim \mathbb{Z}$ it is necessary that we draw first its ending point. Thus, we actually have to decompose a Brownian bridge at its minimum which justifies the conditioning on ω_T at (3.7). Proposition 3.2 provides precisely the method for drawing the minimum and the time instance when it is achieved for a path $\omega \sim \mathbb{Z}$ given its ending point; the following theorem gives the way of filling in the remainder of ω .

We denote by $R(\delta) = \{R_t(\delta); t \in [0, 1]\}$ a 3-dimensional Bessel bridge of unit length from 0 to $\delta \geq 0$ and by $W^c = (W | m = b, \tau = t, W_T = a)$ the Brownian motion of length T starting at 0 and conditioned on obtaining its minimum b at time t and ending at a .

Theorem 3.2 *The processes $\{W_s^c; s \in [0, t]\}$, $\{W_s^c; s \in [t, T]\}$ are independent*

with

$$\begin{aligned} \{W_s^c; s \in [0, t]\} &\stackrel{d}{=} \sqrt{t} \{R_{(t-s)/t}(\delta_1); s \in [0, t]\} + b, \\ \{W_s^c; s \in [t, T]\} &\stackrel{d}{=} \sqrt{T-t} \{R_{(s-t)/(T-t)}(\delta_2); s \in [t, T]\} + b, \end{aligned}$$

where $\delta_1 = -b/\sqrt{t}$ and $\delta_2 = (a - b)/\sqrt{T - t}$.

Proof: This is Proposition 2 of Asmussen et al. (1995) with the difference that we rescaled the Bessel processes to obtain bridges of unit length and the Brownian path to incorporate for the case that decomposition of a path of arbitrary length T is requested. □

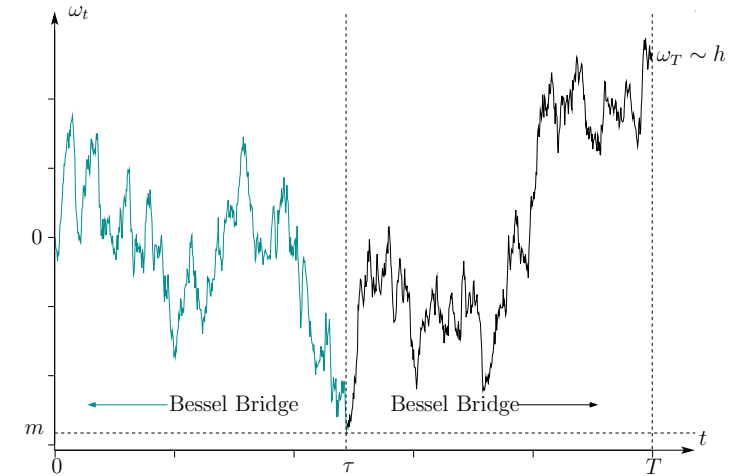


Figure 3.2: The decomposition of $\omega \sim \mathbb{Z}$ at its minimum. The order in the simulation of the involved random elements is ω_T, m, τ . The two independent Bessel bridges connect the already simulated instances of ω .

It is well-known (see for instance Bertoin and Pitman, 1994) that if $W_1^{br}, W_2^{br}, W_3^{br}$ are three independent standard Brownian bridges (bridges of unit length, starting

and ending at 0) then we can produce a Bessel bridge $R(\delta)$ as:

$$R_t(\delta) = \sqrt{(\delta t + W_{1,t}^{br})^2 + (W_{2,t}^{br})^2 + (W_{3,t}^{br})^2}, \quad t \in [0, 1].$$

Recall also (Bertoin and Pitman, 1994) that a standard Brownian bridge W^{br} can be easily expressed in terms of an unconditional Brownian motion W started at 0 via the transformation $W_t^{br} = W_t - t W_1$.

3.3.2 The extension of the Exact Algorithm

Theorem 3.2 can be used to generate a subroutine, say `decompose`($T, \beta_0, \beta_T, t, b, \mathbf{s}$), that returns the location at the time instances $\mathbf{s} = \{s_1, \dots, s_n\}$ of a Brownian path on $[0, T]$ started at β_0 , conditioned on obtaining its minimum b at time t and finishing at β_T . We can now extend EA1 to simulate from (3.2) when the functional $\phi \geq 0$ in (3.4) is not necessarily bounded above but satisfies the weaker condition $\limsup_{u \rightarrow \infty} \phi(u) < \infty$. In translating the construction to the present setting, the Poisson process Φ defined in Theorem 3.1 in this case is a point process on rectangles whose height depends on the current proposed path ω .

Exact Algorithm 2 (EA2)

1. Initiate a path $\omega \sim \mathbb{Z}$ on $[0, T]$ by drawing $\omega_T \sim h$.
 2. Simulate its minimum m and the moment when it is achieved τ .
 3. Calculate $r(m)$.
 4. Produce a realisation $\{(\psi_1, v_1), (\psi_2, v_2), \dots, (\psi_\kappa, v_\kappa)\}$ of Φ in $[0, T] \times [0, r(m)]$.
 5. Call `decompose()` to construct ω at $\{\psi_1, \psi_2, \dots, \psi_\kappa\}$.
 6. Evaluate N .
 7. If $N = 0$ goto (8), else goto (1).
 8. Output the currently constructed skeleton $S(\omega)$ of ω .
-

Note that, in comparison with the skeleton returned by the initial algorithm of Section 3.2.1, this time the skeleton $S(\omega)$ is more informative about the underlying continuous path ω since it includes its minimum and the instance when it is achieved. As with the initial algorithm, we can continue filling in the accepted skeleton $S(\omega)$ at any other time instances under the two independent Bessel bridges dynamics.

3.4 The efficiency of the Exact Algorithm

It is critical for the efficiency of the Exact Algorithm that we can exploit the Markov property of the target diffusion (3.2) and merge skeletons of some appropriately chosen length T (or even of varying lengths) to construct a complete skeleton on some larger interval. In what follows we preserve T to denote the length of the time interval upon which the Exact Algorithm is applied (and which is typically smaller than the length, say l , of the sample path of interest).

Assume that we apply the Exact Algorithm on X over the time interval $[0, T]$ for some starting point $x \in \mathbf{R}$. Let ϵ denote the probability of accepting a proposed path and D the number of the Poisson process points needed to decide whether to accept the proposed path. Let $N(T)$ be the total number of Poisson process points needed until the first accepted path.

Proposition 3.3 *Consider EA1 with $\phi \leq r$. Then,*

$$\epsilon \geq \exp(-r \cdot T), \quad \mathbb{E}[D] = r \times T.$$

For EA2 with $r = r(m)$ defined as in (3.6):

$$\epsilon \geq \exp\{-\mathbb{E}_{\mathbb{Z}}[r(m)] \cdot T\}, \quad \mathbb{E}[D] = \mathbb{E}_{\mathbb{Z}}[r(m)] \times T.$$

In both cases:

$$\mathbb{E}[N(T)] \leq \mathbb{E}[D] / \epsilon.$$

Proof: In the Appendix.

It is natural to ask how we should implement the algorithms in order to simulate the diffusion on the time interval $[0, KT]$ for some positive integer K . For concreteness, consider EA1. Proposition 3.3 suggests that implementing the rejection sampling algorithm on the entire interval will incur a computational cost which is $O(rKTe^{rKT})$, which will be huge even for moderate K values. On the other hand the simulation problem can be broken down into K simulations performed sequentially, each performed on an interval of length T , and each incurring a cost which is $O(rTe^{rT})$ and therefore giving an overall cost which is $O(rKTe^{rT})$, i.e. linear in the length of the time interval required.

Clearly there are important practical problems of algorithm optimisation (by choosing T) for simulating from a diffusion on a fixed time scale. We shall not address these here, except to remark that the larger r is, the smaller T ought to be.

It is difficult to obtain explicit results on the efficiency of EA2. From Proposition 3.3 it is clear that to show that the algorithm runs in finite time we need to bound $\mathbb{E}_{\mathbb{Z}}[r(m)]$ which is difficult. The following Proposition gives a general result we have obtained, and which will be found useful in the case of the logistic growth model example of Section 3.5.

Proposition 3.4 *Assume that the drift function α is bounded below and there exists $k > 0$ and b_o such that $r(b) \leq \exp(-kb)$ for $b \leq b_o$. Then for any starting value x and any length T of the time interval under consideration $\mathbb{E}_{\mathbb{Z}}[r(m)] < \infty$. Therefore for EA2, the expected number of proposed Poisson points needed to accept a sample path is finite.*

Proof: In the Appendix.

Note that once we have established a theoretical result about the finiteness of $\mathbb{E}_{\mathbb{Z}}[r(m)]$ we can easily proceed to a Monte Carlo estimation of its actual value.

3.4.1 User-impatience bias

Many non-trivial simulation schemes return results in unbounded, random times. If the experiment is designed in a way that its running time is *not* independent of the output, then the results can be “user-impatience” biased; the experimenter might be tempted (or even obliged) to stop a long run and re-start the simulation procedure. So, the sample will be biased toward the observations that require shorter runs. The case when this situation has appeared in its most concrete form is the Coupling From the Past (CFTP) technique for the perfect simulation of distributions using Markov chains. The discussion in section 7 of Propp and Wilson (1998a) presents analytically the problem in the CFTP case.

A similar problem can be identified for EA2. The time to decide about the acceptance or otherwise of the current proposed path ω depends on ω . If D is the number of Poisson process points to decide about a proposed path, then $\mathbb{E}[D|\omega] = T \times r(m(\omega))$. Since, in theory, the proposed path ω can produce any small value, it is impossible to bound $\mathbb{E}[D|\omega]$ uniformly over all ω . For the example with the logistic growth model we present in Section 3.5 we had to adjust the length T of the time interval to which EA2 is applied, according to the values of the parameters. In this way we have not only approximately optimised the efficiency, but also minimised the effects of the user-impatience bias.

In line with the suggestions in Propp and Wilson (1998a), the experimenter should try some preliminary executions of the algorithm to get some evidence about the distribution of the running time and, if necessary (and possible), re-design the algorithm so that this distribution is of small variance.

3.5 Application: the logistic growth model

EA2 can be applied to the stochastic analogue of the logistic growth model:

$$dV_t = R V_t (1 - V_t/C) dt + \beta V_t dW_t, \quad t \in [0, T], \quad V_0 = v > 0, \quad (3.8)$$

for positive parameters R, β, C . This diffusion is useful for modelling the growth of populations. The instantaneous population of some species V_t grows, in the absence of any restraints, exponentially fast in t with growth rate per individual equal to R . The actual evolution of the population is cut back by the saturation inducing term $(1 - V_t/C)$. The constant $C > 0$ is called the carrying capacity of the environment and usually represents the maximum population that can be supported by the resources of the environment. The parameter β represents the effect of the noise on the dynamics of V . For an example of how (3.8) can be derived see Goel and Richter-Dyn (1974), ch.6.

After setting $X_t = -\log(V_t)/\beta$ we obtain the SDE:

$$dX_t = \left\{ \frac{\beta}{2} - \frac{R}{\beta} + \frac{R}{\beta C} \exp(-\beta X_t) \right\} dt + dW_t, \quad X_0 = x = -\frac{\log(v)}{\beta}. \quad (3.9)$$

It can be shown using the theory of scale and speed densities, see ch.15, section 7 of Karlin and Taylor (1981), that w.p.1 the original process V does not hit 0 in finite time for any values of the parameters, so considering the $\log(V_t)$ is valid. Let α be the drift function of the modified SDE (3.9). Then:

$$(\alpha^2 + \alpha')(u) = \frac{R^2}{\beta^2 C^2} \exp(-2\beta u) - \frac{2R^2}{\beta^2 C} \exp(-\beta u) + \left(\frac{\beta}{2} - \frac{R}{\beta}\right)^2, \quad u \in \mathbf{R}.$$

This function is bounded below and satisfies $\limsup_{u \rightarrow +\infty} (\alpha^2 + \alpha')(u) < \infty$, so it satisfies the conditions required for the applicability of EA2.

We follow the analysis of Section 3.2 to construct the Radon-Nikodym derivative (3.4) that is appropriate for rejection sampling. The ending point of the biased

Brownian motion $\mathbb{Z}$ must be distributed according to a density $h(u)$ proportional to:

$$\exp \left\{ \left(\frac{\beta}{2} - \frac{R}{\beta} \right) u - \frac{R}{\beta^2 C} e^{-\beta u} - \frac{(u-x)^2}{2T} \right\} \propto \exp \left\{ -\frac{(u-g_1)^2}{2T} - g_2 e^{-\beta u} \right\} \quad (3.10)$$

for $g_1 = x + T(\beta/2 - R/\beta)$, $g_2 = R/(\beta^2 C)$. We can draw from this density in an efficient way via rejection sampling with envelope function

$$\exp \left\{ -\frac{(u-g_1)^2}{2T} - g_2 e^{-\beta g_1} (1 - \beta(u-g_1)) \right\}, \quad u \in \mathbf{R}.$$

This function is proportional to the density of a $\mathcal{N}(g_1 + T\beta g_2 e^{-\beta g_1}, T)$. We can now obtain (3.4) for $\mathbb{Z}$ the probability measure that corresponds to the biased Brownian motion with ending point distributed according to (3.10) and:

$$\phi(u) = \frac{R^2}{2\beta^2 C^2} e^{-2\beta u} - \frac{R^2}{\beta^2 C} e^{-\beta u} + \frac{R^2}{2\beta^2} \geq 0, \quad u \in \mathbf{R}.$$

If m is the minimum of a proposed path $\omega \sim \mathbb{Z}$ we can use (3.6) to find an upper bound for the mapping $t \mapsto \phi(\omega_t)$:

$$r(\omega) \equiv r(m) = \max\{\phi(m), R^2/2\beta^2\}.$$

It is now easy to apply the algorithm. Empirical evidence suggests that the efficiency of EA2 for this problem is sensitive to R, β and, the ratio v/C . Table 3.1 gives a comparison of running times as parameters vary. Since we can take advantage of the Markov property of the process V to merge exact skeletons of some appropriate length T to reach the instance $l = 10$, we have approximately optimised over the choice of T in each case. We also compare the results with the Euler scheme that approximates (3.8) with the discrete time process:

$$Y_0 = v, \quad Y_{(i+1)h} = Y_{ih} + R Y_{ih} (1 - Y_{ih}/C) h + \beta Y_{ih} \sqrt{h} Z_{i+1},$$

where Z_{i+1} , $i \geq 0$, are iid with $Z_1 \sim \mathcal{N}(0,1)$. We applied the Euler scheme for increments of the type $h = 2^{-n}$, for $n = 1, 2, \dots$. To find the increment h that generates a good approximation to the true process, we carried out four separate Kolmogorov-Smirnov tests to pairs of samples of size 10^5 from X_t, Y_t , for $t = 1/2, 1, 5, 10$ (all eight samples were independent) and found the maximum h for which at least three of the four KS tests did not reject the null hypothesis of equivalence in distribution of the two samples at a 10% significance level. This ensures a significance level for the composite test of 0.052.

From the evidence presented in Table 3.1 and other simulations we have performed, EA2 for the logistic growth model seems to be generally computationally quicker than the Euler approximation. However when R/β is very large, EA2 becomes rather slow. For $R = 1$ and $\beta = 0.1$ for instance the “sufficiently accurate” Euler approximation appears to be considerably more rapid, although of course Euler still only produces an approximation to the true distribution.

To illustrate the output obtained from the algorithm, in Fig.3.3 we show an exact skeleton of the target diffusion $\{V_t; t \in [0, 1000]\}$, $V_0 = 1000$, for parameter values $C = 1000$, $R = 0.5$, and $\beta = 0.5$ as generated by EA2. We applied EA2 at time intervals of length $T = 0.5$ and we show the instances of the skeleton for the times $t = 0, 1, 2, \dots$. We will use these data in Section 3.6.1.

3.6 Applications of the Exact Algorithm

3.6.1 Monte Carlo maximum likelihood for discretely observed diffusions

An important problem in inference from diffusions involves the calculation of maximum likelihood estimators from a finite discrete set of observations of the diffusions. Assume that the drift b and the diffusion coefficient σ of the SDE in (3.1) have parametric forms which depend on certain unknown parameters θ . As a

parameters			Exact				Euler	
v	R	β	T	$\bar{D}$	$\bar{I}$	time	h	time
1,000	0.01	0.10	5.00	0.0245	1.0011	1.2 sec	2^{-5}	19.4 sec
50	0.01	0.10	5.00	0.0250	1.0228	1.9 sec	2^{-3}	7.8 sec
1,800	0.01	0.10	5.00	0.0367	1.0173	1.4 sec	2^{-6}	35.2 sec
1,000	1.00	1.00	0.25	0.1623	1.0652	21.6 sec	2^{-9}	256.9 sec
1	1.00	1.00	0.25	0.1273	1.1174	22.3 sec	2^{-8}	137.9 sec
3,500	1.00	1.00	0.25	0.2396	1.0808	25.7 sec	2^{-10}	558.1 sec
1,000	1.00	0.10	0.10	5.0031	1.0223	284.2 sec	2^{-6}	35.4 sec
750	1.00	0.10	0.10	5.0009	1.0458	291.7 sec	2^{-7}	86.4 sec
1,250	1.00	0.10	0.10	5.0000	1.0398	297.6 sec	2^{-8}	152.9 sec

Table 3.1: A comparative presentation of EA2 and the Euler scheme as applied to the logistic growth model for different parameter values. In each case, simulation over the time interval $[0, 10]$ was performed. $\bar{D}$ denotes the mean number of Poisson process points needed to decide about a proposed path; $\bar{I}$ denotes the mean number of proposed paths until (and including) the first successful one; and T denotes the time interval used to apply EA2. The times given represent the overall running time to obtain a sample of 100,000 paths. The Euler simulations are given for sufficiently small h to satisfy the composite KS test described in the main text.

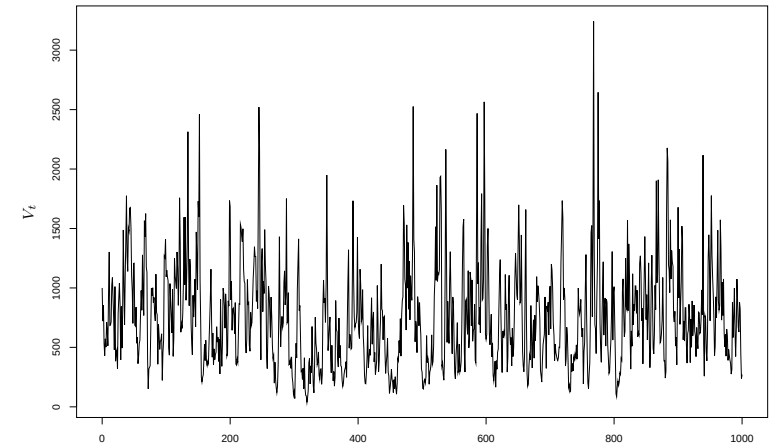


Figure 3.3: An exact skeleton of length 1000 of the logistic growth model (3.8), for $R = 0.5$, $\beta = 0.5$ and initial value $v = 1000$.

running example in this section, consider the logistic growth model (3.8), where $\theta = (R, C, \beta)$. The goal is to find the MLE of θ based on discrete-time observations

of X , say $X_{t_0} = x_0, X_{t_1} = x_1, \dots, X_{t_n} = x_n$. It is common practise to set $t_0 = 0$ and treat the first point x_0 as fixed, although this is not essential. Let $p_t(x, y; \theta)$ be the transition density of the diffusion process, i.e. the Lebesgue density of the transition measure $P[X_t \in dy | X_0 = x; \theta]$. The likelihood function is the density of the observed data $x_1, \dots, x_n$ (conditionally on $X_{t_0} = x_0$) as a function of θ ,

$$\mathcal{L}(\theta) = \prod_{i=1}^n p_{\Delta t_i}(x_{i-1}, x_i; \theta) = \prod_{i=1}^n L_i(\theta), \quad \Delta t_i = t_i - t_{i-1}, \quad (3.11)$$

where we define

$$L_i(\theta) = p_{\Delta t_i}(x_{i-1}, x_i; \theta), \quad i = 1, \dots, n,$$

to be the likelihood contribution of the i -th data point. Notice that the factorisation in (3.11) is due to the Markov property. The aim is to maximise (3.11) as a function of θ , however this is a daunting task since the transition density $p_t(x, y; \theta)$ is typically not analytically available. Hence, this problem has attracted considerable attention and constitutes a very active area of research; a recent review and references can be found in Sørensen (2004).

Here we present a simple, fast and efficient Monte Carlo method for estimating $L_i(\theta)$, $i = 1, \dots, n$, for any given value θ . The MLE can then be found by a grid search method, although more sophisticated stochastic search algorithms can also be applied for its identification. Our approach is applicable whenever EA1 or EA2 can simulate from the diffusion process which generates the data (or at least from the transformed to have unit diffusion coefficient process). The Exact Algorithm together with a variance reduction technique will provide independent unbiased estimators $\hat{L}_1(\theta), \dots, \hat{L}_n(\theta)$. We proceed by describing how to find $\hat{L}_1(\theta)$, since the derivation for any other $i > 1$ proceeds in the same way.

Let $\tilde{p}_t(x, y; \theta)$ be the transition density of the transformed to unit diffusion coefficient process (3.2), that is of the process obtain by the transformation $X_t \mapsto$

$\eta(X_t)$, where η is defined in (3.3). Then,

$$p_t(x, y; \theta) = \tilde{p}_t(\eta(x), \eta(y); \theta) |\eta'(y)|,$$

where η might also depend on θ (as for example for the logistic growth model). Let S be a skeleton of the transformed process started at $\eta(X_0) = \eta(x)$, returned by our Exact Algorithm on $[0, t + \gamma]$, $\gamma > 0$, for $t \equiv t_1$ the instance of the first datum. We can write,

$$S = \{(u_0, \eta(X_{u_0})), \dots, (u_l, \eta(X_{u_l}))\},$$

where $u_0 = 0, u_l = t + \gamma$ and $\eta(X_{u_0}) = \eta(x)$, but otherwise the time instances u_i , $0 < i < l$, will be random. We define,

$$\tilde{p}_t(z, w | \cdot; \theta) = P[\eta(X_t) \in dw | \eta(X_0) = z, \cdot; \theta] / dw,$$

to be the density of $\eta(X_t)$ given the starting value and any other random elements. Then, by the standard property of conditional expectation the following is true,

$$\tilde{p}_t(\eta(x), \eta(y); \theta) = \begin{cases} \mathbb{E}[\tilde{p}_t(\eta(x), \eta(y) | S; \theta)], & \text{if } \phi \text{ is bounded} \\ \mathbb{E}[\tilde{p}_t(\eta(x), \eta(y) | S, m, \tau; \theta)], & \text{if } \limsup_{u \rightarrow \infty} \phi(u) < \infty \end{cases}$$

where m is the minimum of the diffusion path on $[0, t + \gamma]$ and τ is the time when the minimum occurs. In the second case, both m and τ are elements of S , see the discussion at the end of Section 3.3.2, but we write them separately to distinguish between the skeleton produced by EA1 and EA2.

We first consider the case of EA1, when S is simulated as in Section 3.2.1. This is the case when the functional ϕ is bounded. Notice that due to the Markov

property,

$$\tilde{p}_t(\eta(x), \eta(y) \mid S; \theta) \equiv \tilde{p}_t(\eta(x), \eta(y) \mid \eta(X_{t_-}), \eta(X_{t_+}); \theta), \quad (3.12)$$

where we define

$$t_- \equiv \sup\{u_i; i = 0, \dots, l : u_i < t\}, \quad \text{and} \quad t_+ \equiv \inf\{u_i; i = 0, \dots, l : u_i > t\}.$$

By construction, both t_- and t_+ are well defined, since the skeleton contains at least the instances $u_0 \equiv 0 < t$, and $u_l \equiv t + \gamma > t$. However, we have already argued (Section 3.2.1) that, given the skeleton, the distribution of the process at any time $s \in (0, t + \gamma)$ is given by the law of the Brownian bridge between the points of the skeleton adjacent to s . Therefore, (3.12) becomes

$$\frac{1}{\sqrt{2\pi\delta}} \exp \left\{ -\frac{1}{2\delta} \left(\eta(y) - \eta(X_{t_-}) - \frac{t - t_-}{t_+ - t_-} (\eta(X_{t_+}) - \eta(X_{t_-})) \right)^2 \right\}, \quad (3.13)$$

for $\delta = (t - t_-)(t_+ - t)/(t_+ - t_-)$. Note that the expression (3.13) does not depend on θ .

We turn now to EA2. When S is simulated as in Section 3.3, where we first simulate the minimum of the diffusion paths, we have that

$$\tilde{p}_t(\eta(x), \eta(y) \mid S, m, \tau; \theta) \equiv \tilde{p}_t(\eta(x), \eta(y) \mid \eta(X_{t_-}), \eta(X_{t_+}), m, \tau; \theta). \quad (3.14)$$

Let $q(t, z, w)$ be the transition density of the 3-dimensional Bessel process, which is known analytically and can be readily computed (see for example pg.446 of Revuz and Yor, 1991). Then, using Theorem 3.2 and working from first principles, it can

be shown that (3.14) is,

$$\begin{cases} \frac{q\left(\frac{t-t_-}{t_+-\tau}, \frac{\eta(X_{t_-})-m}{\sqrt{t_+-\tau}}, \frac{\eta(X_{t_+})-m}{\sqrt{t_+-\tau}}\right) \times q\left(\frac{t_+-t}{t_+-\tau}, \frac{\eta(X_t)-m}{\sqrt{t_+-\tau}}, \frac{\eta(X_{t_+})-m}{\sqrt{t_+-\tau}}\right)}{q\left(\frac{t_+-t_-}{t_+-\tau}, \frac{\eta(X_{t_-})-m}{\sqrt{t_+-\tau}}, \frac{\eta(X_{t_+})-m}{\sqrt{t_+-\tau}}\right) \sqrt{t_+-\tau}}, & \tau < t \\ \frac{q\left(\frac{t_+-t}{\tau}, \frac{\eta(X_{t_+})-m}{\sqrt{\tau}}, \frac{\eta(X_t)-m}{\sqrt{\tau}}\right) \times q\left(\frac{t-t_-}{\tau}, \frac{\eta(X_t)-m}{\sqrt{\tau}}, \frac{\eta(X_{t_-})-m}{\sqrt{\tau}}\right)}{q\left(\frac{t_+-t_-}{\tau}, \frac{\eta(X_{t_+})-m}{\sqrt{\tau}}, \frac{\eta(X_{t_-})-m}{\sqrt{\tau}}\right) \sqrt{\tau}}, & \tau \geq t \end{cases} \quad (3.15)$$

which as in (3.13) does not depend on θ .

Therefore, we can now construct a Monte Carlo algorithm for the unbiased estimation of $L_1(\theta)$, for any given value of θ . We describe it analytically for the more involved situation where S is simulated as in Section 3.3; it is clear how to modify the algorithm for the simpler case when ϕ is bounded.

Monte Carlo estimation of $L_1(\theta)$

1. For $j = 1 : k$, repeat:
 2. Draw a skeleton S^j on $[0, t_1 + \gamma]$ started from $\eta(X_{t_0}) = \eta(x_0)$. Let (τ^j, m^j) be the time and value of the minimum of S^j .
 3. Compute $\tilde{p}_{t_1}(\eta(x_0), \eta(x_1) \mid S^j, m^j, \tau^j)$ according to (3.15).
 4. Goto (1).
 5. Output $\hat{L}_1(\theta) = \frac{1}{k} \sum_{j=1}^k \tilde{p}_{t_1}(\eta(x_0), \eta(x_1) \mid S^j, m^j, \tau^j) |\eta'(x_1)|$.
-

We have deliberately omitted θ from the expression of the transition density conditionally on the skeleton elements since this conditioned density does not depend on θ .

As an initial illustration of the validity of the algorithm we implemented it for the case of the Brownian motion with drift $X_t = at + \sigma W_t$ with parameters $\theta = (a, \sigma)$ when the likelihood over a set of data is explicitly known. We generated a data set $\{X_0, X_1, \dots, X_{1000}\}$, $X_0 = 0$, after using $a = 1$ and $\sigma = 1$ and applied the algorithm that uses the Bessel process described above (although we could have used the simpler algorithm with the Brownian bridges). For this trivial example

the Exact Algorithm will propose paths from a probability measure that coincides with the probability measure of the target process X , so they will be accepted w.p.1 and the skeletons S^j will always include precisely three points obtained at the instances $u_0 = 0$, $u_1 = \tau$ and $u_2 = t + \gamma$. In our case, $t = 1$; also, we chose $\gamma = 0.5$. Fig.3.4 shows the logarithm of the actual profile likelihoods for the two parameters and their estimation as provided by our algorithm for 5000 Monte Carlo iterations and a grid (for both parameters) of increment 0.05.

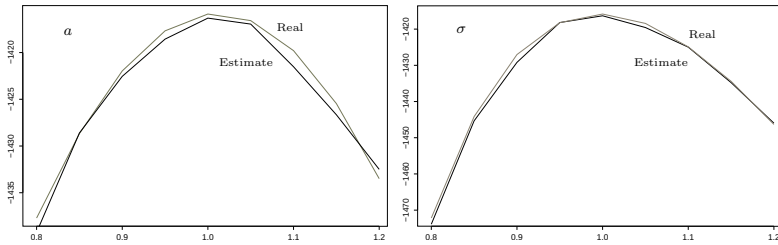


Figure 3.4: The actual and the estimated (based on 5000 samples) profile log-likelihoods for the parameters a , σ in the case of the Brownian motion with drift.

We then applied our algorithm to the logistic growth model (3.8) and for the data shown in Fig.3.3 which are generated for $C = 1000$, $R = 0.5$, $\beta = 0.5$ and initial value $v = 1000$. The data are equidistant with time increment 1 and are obtained on the time interval $[0, 1000]$. We considered C as known and tried to find the MLE for (R, β) . Fig.3.5 shows the logarithm of the profile likelihoods for both parameters as estimated by our algorithm for 10,000 Monte Carlo iterations and grid increment 0.05; we used $\gamma = 0.1$.

The success of our Monte Carlo method lies in the fact that both $\tilde{p}_t(z, w | S)$ and $\tilde{p}_t(z, w | S, m, \tau)$ have known explicit forms, given in (3.13) and (3.15) respectively. Thus the only source of Monte Carlo error is on the averaging over the skeletons. The technique of obtaining closed form expressions when conditioning on certain random variables, and then averaging over these variables using Monte Carlo is

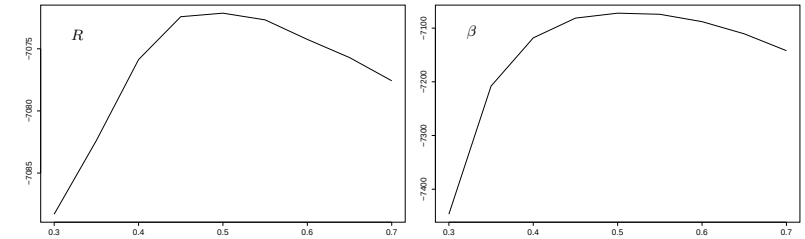


Figure 3.5: The estimated (based on 10,000 samples) profile log-likelihoods for the parameters R , β of the logistic growth model.

known as *Rao-Blackwellisation* (see for example Gelfand and Smith, 1990) and is a well-established variance reduction method. In this paper we have only briefly outlined the potential of this method, and it will be pursued in detail in subsequent work.

3.6.2 Exact simulation of conditioned diffusions

In this section, we address the problem of exact simulation from SDEs of the type (3.2) conditioned on $X_T = y$ for a fixed $y \in \mathbf{R}$. This is an important problem, in particular for problems of Bayesian inference for discretely observed diffusions. We will see that our methods extend easily to this case. However, first we demonstrate why conventional simulation methods are problematic for conditioned diffusions.

If Y is a diffusion with the law of (3.2) conditioned on $X_T = y$, then it is well known (for instance by h -transforms, see for example Rogers and Williams, 2000, section IV.39), that Y satisfies the SDE:

$$dY_t = dW_t + \left(\alpha(Y_t) + \frac{\partial \log h(T-t, Y_t)}{\partial z} \right) dt, \quad (3.16)$$

where $h(s, z)$ is the density of the diffusion (3.2) started from z at location y and

time s , i.e. $P_z(X_s \in dy)/dy$. It satisfies the parabolic partial differential equation

$$\frac{\partial h(s, z)}{\partial s} = \frac{1}{2} \frac{\partial^2 h(s, z)}{\partial z^2} + \alpha(z) \frac{\partial h(s, z)}{\partial z}. \quad (3.17)$$

Conventional approximate methods for simulating from (3.16) pose great difficulties. They require a preliminary numerical calculation of h from (3.17) and have to deal with the fact that the drift explodes as t approaches T .

In contrast, the Radon-Nikodym derivative between the law of Y and that of a Brownian bridge from $(0, x)$ to (T, y) is proportional to

$$\exp \left\{ - \int_0^T \phi(\omega_t) dt \right\},$$

for ϕ as defined in (3.5). So, the Exact Algorithm can easily be adapted to conditional simulation simply by omitting the biased Brownian motion step (since ω_T is now considered fixed). It will be feasible under Conditions 1,3 given in Section 3.2 and the property that $\limsup_{u \rightarrow \infty} \phi(u)$ and $\limsup_{u \rightarrow -\infty} \phi(u)$ are not both $+\infty$.

3.7 Conclusions

This paper has introduced an algorithm for perfect simulation of diffusion processes, relaxing some of the regularity conditions imposed by the previous method of Beskos and Roberts (2005) [Article 2]. The algorithm seems to be computationally efficient in many cases as our simulation study for the logistic growth model demonstrated. Currently, we know little about why the method's efficiency relative to a standard Euler-scheme alternative varies so considerably, and a systematic simulation study and theoretical analysis will be required to investigate this further.

We outline two potential application areas, both which warrant further investigation. The Monte Carlo Maximum Likelihood approach capitalises on the

form of the output given by EA1 or EA2 to provide unbiased likelihood estimates. Computational efficiency of this and related techniques will rely on being able to provide likelihood estimates which are smooth as a function of the parameter, and with variances which are robust to data sample sizes.

Simulation of conditioned diffusions is a straightforward by-product of EA1 or EA2. However it has a number of potentially important examples. Our primary motivation comes from the need to carry out imputation of unobserved sample paths as part of data-augmentation schemes for Bayesian inference for diffusions.

The major unsolved problem in this area involves the simulation of diffusions whose drift is unbounded in both tails. This is considerably more challenging than the case considered here, mainly due to the difficulties with obtaining simple closed form solutions to two-sided boundary crossing problems for Brownian motion. Extensions to a multivariate context are clearly possible, though they would rely on a multivariate analogue of (3.3), which is not possible in total generality.

APPENDIX

Proof of Proposition 3.1:

Let $\mathcal{C}_T$ be the σ -algebra corresponding to the information about the ending point of the paths. The property $(M | M_T = \rho) \stackrel{d}{=} (N | N_T = \rho)$, for all $\rho \in \mathbf{R}$, can be expressed in a rigorous way as:

$$\mathbb{M}[A | \mathcal{C}_T] = \mathbb{N}[A | \mathcal{C}_T], \quad \mathbb{N} - \text{a.s.},$$

for any $A \in \mathcal{C}$. It suffices to prove that:

$$\mathbb{M}[A] = \mathbb{E}_{\mathbb{N}} \left[\mathbb{I}_A \cdot \frac{f_M}{f_N}(\omega_T) \right], \quad \text{for any } A \in \mathcal{C},$$

Simple calculations give:

$$\begin{aligned}\mathbb{E}_{\mathbb{N}}[\mathbb{I}_A \cdot \frac{f_M}{f_N}(\omega_T)] &= \mathbb{E}_{\mathbb{N}}[\mathbb{E}_{\mathbb{N}}[\mathbb{I}_A \cdot \frac{f_M}{f_N}(\omega_T) | \mathcal{C}_T]] = \\ &= \mathbb{E}_{\mathbb{N}}[\frac{f_M}{f_N}(\omega_T) \cdot \mathbb{N}[A | \mathcal{C}_T]] = \mathbb{E}_{\mathbb{M}}[\mathbb{N}[A | \mathcal{C}_T]] = \\ &= \mathbb{E}_{\mathbb{M}}[\mathbb{M}[A | \mathcal{C}_T]] = \mathbb{M}[A] .\end{aligned}$$

Proof of Proposition 3.2:

From pg.95 of Karatzas and Shreve (1991) we obtain:

$$\mathbb{P}[m \in db | \omega_T = a] = \frac{2}{T}(a - 2b) \exp\left\{-\frac{(a - 2b)^2}{2T} + \frac{a^2}{2T}\right\} db, \quad b \leq \min\{a, 0\} .$$

We can easily derive the representation:

$$[m | \omega_T = a] \stackrel{d}{=} (a - \sqrt{2TE + a^2})/2 \equiv Z_1 .$$

It is clear from (3.7) that:

$$\mathbb{P}[\tau \in dt | m = b, \omega_T = a] \propto \frac{1}{\sqrt{t^3(T-t)^3}} \exp\left\{-\frac{b^2}{2t} - \frac{(a-b)^2}{2(T-t)}\right\} dt, \quad t \in [0, T] .$$

If $V := (T - \tau)/\tau$ and f denotes the density of V conditionally on $\omega_T = a$, $m = b$ then:

$$f(y) \propto y^{-3/2} \exp(-\frac{c_1}{y} - c_2 y) + y^{-1/2} \exp(-\frac{c_1}{y} - c_2 y), \quad y \geq 0 , \quad (3.18)$$

for $c_1 = (a - b)^2/2T$, $c_2 = b^2/2T$. This density can be identified as the mixture of an $IGau(\sqrt{c_1/c_2}, 2c_1)$ and an $1/IGau(\sqrt{c_2/c_1}, 2c_2)$. The ratio of the integral of

the left summand in (3.18) to the integral of both summands is:

$$p = \frac{1}{1 + \sqrt{\frac{c_1}{c_2}}} ,$$

therefore, w.p. p we draw $I_1 \sim IGau(\sqrt{c_1/c_2}, 2c_1)$ and w.p. $(1 - p)$ we draw $I_2 \sim 1/IGau(\sqrt{c_2/c_1}, 2c_2)$. If we use a $U \sim \text{Un}(0, 1)$ to decide for the choice between the densities of the mixture we can obtain that for the Z_2 defined in Proposition 3.2, $[Z_2 | Z_1 = b] \stackrel{d}{=} [\tau | m = b, \omega_T = a]$.

Proof of Proposition 3.3:

When $\phi \leq r$, it is clear from (3.4) that $\epsilon \geq \exp(-r \cdot T)$. Trivially, Theorem 3.1 gives that $\mathbb{E}[D] = r \times T$.

When $\limsup_{u \rightarrow \infty} \phi(u) < \infty$ and $r(m) = \sup\{\phi(u); u \geq m\}$, then we can bound ϵ using Jensen's inequality:

$$\epsilon = \mathbb{E}_{\mathbb{Z}}\left[\exp\left\{-\int_0^T \phi(\omega_t) dt\right\}\right] \geq \mathbb{E}_{\mathbb{Z}}[\exp\{-r(m) \cdot T\}] \geq \exp\{-\mathbb{E}_{\mathbb{Z}}[r(m)] \cdot T\} .$$

Also, $\mathbb{E}[D] = \mathbb{E}_{\mathbb{Z}}[\mathbb{E}[D | m]] = \mathbb{E}_{\mathbb{Z}}[r(m)] \times T$.

For both cases, to estimate $N(T)$, assume that $\{\omega_1, \omega_2, \dots\}$ is a sequence of proposed paths, D_i is the number of Poisson process points needed to decide about the acceptance or the rejection of the i -th proposed path for any $i \geq 1$, and $I = \inf\{i \geq 1 : \omega_i \text{ is accepted}\}$. Assume also that $\mathbb{E}[D | A]$, $\mathbb{E}[D | A^c]$ are the expected number of Poisson process points conditionally on accepting and rejecting the proposed path respectively. Then, if $\mathbb{E}[D] < \infty$:

$$\begin{aligned}\mathbb{E}[N(T)] &= \mathbb{E}[D_1 + D_2 + \dots + D_I] = \mathbb{E}[\mathbb{E}[D_1 + D_2 + \dots + D_I | I]] = \\ &= \mathbb{E}[(I - 1) \cdot \mathbb{E}[D | A^c] + \mathbb{E}[D | A]] = \\ &= \left(\frac{1}{\epsilon} - 1\right) \mathbb{E}[D | A^c] + \mathbb{E}[D | A] = \frac{\mathbb{E}[D]}{\epsilon} .\end{aligned}$$

If $\mathbb{E}[D] = \infty$ then $\mathbb{E}[N(T)] = \infty$.

Proof of Proposition 3.4:

$\mathbb{E}^x[\cdot]$ will denote expectations of path elements conditionally on the starting point of the path being $x \in \mathbf{R}$. $\mathbb{E}_{\mathbb{Z}}^x[r(m)]$ is decreasing in x ; the minimum of the $\mathbb{Z}$ -path is stochastically increasing in x and $r(m) = \sup\{\phi(u); u \geq m\}$ is decreasing. To simplify the proof, we can choose $x < b_o$; for this x , $r(m) \leq \exp(-km)$ for any m in its domain $(-\infty, x)$. Consider a real δ such that $\alpha(u) \geq \delta$, $u \in \mathbf{R}$. It can be easily shown that:

$$\text{biased Brownian motion } \mathbb{Z} \stackrel{\text{stoch.}}{\geq} \text{Brownian motion (BM) with drift } \delta, \quad (3.19)$$

where both processes are considered on $[0, T]$ and for the same starting point x . This follows after representing the BM with drift δ as a biased BM with ending point distributed according to $f(u) \propto \exp\{\delta u - (u - x)^2/2T\}$. Recall that the ending point of $\mathbb{Z}$ is distributed according to $h(u) \propto \exp\{A(u) - (u - x)^2/2T\}$. Also:

$$\left\{\log\left(\frac{h}{f}\right)\right\}'(u) = \alpha(u) - \delta \geq 0,$$

so h/f is increasing. That indicates that the distribution with density h is stochastically larger than that with density f , which implies (3.19).

We denote by $\mathbb{W}_\delta$ the law of the Brownian motion with drift δ . Property (3.19) yields:

$$\mathbb{E}_{\mathbb{Z}}[r(m)] \leq \mathbb{E}_{\mathbb{W}_\delta}[r(m)] \leq \mathbb{E}_{\mathbb{W}_\delta}[\exp(-km)].$$

Using Girsanov theorem we can derive the density of the minimum of the Brownian motion with drift (see for instance Borodin and Salminen, 2002):

$$\mathbb{P}[m \in db]/db = 2\mathcal{N}_{x+\delta T, T}(b) + 2\delta \cdot \exp\{2\delta(b-x)\} \cdot \Phi\left(\frac{b-x+\delta T}{\sqrt{T}}\right), \quad b \leq x,$$

where $\Phi(u) = \int_{-\infty}^u \mathcal{N}_{0,1}(y)dy$, $u \in \mathbf{R}$. It is easy now to check that this density has finite exponential expectations, i.e. $\mathbb{E}_{\mathbb{W}_\delta}[\exp(-km)] < \infty$.

Article 4

Exact and Computationally Efficient Likelihood–Based Estimation for Discretely Observed Diffusion Processes

4.1 Introduction

Diffusion processes are extensively used for modelling continuous-time phenomena in many scientific areas; an incomplete list with some indicative references includes economics (Black and Scholes, 1973; Chan et al., 1992; Cox et al., 1985; Merton, 1971), biology (McAdams and Arkin, 1997), genetics (Kimura and Ohta, 1971; Shiga, 1985), chemistry (Gillespie, 1976, 1977), physics (Obukhov, 1959) and engineering (Pardoux and Pignol, 1984). Their appeal lies in the fact that the model is built by specifying the instantaneous mean and variance of the process through an SDE. Specifically, a diffusion process V is defined as the solution of an SDE of the type:

$$dV_s = b(V_s; \theta) ds + \sigma(V_s; \theta) dW_s, \quad (4.1)$$

driven by the scalar Brownian motion W . The functionals $b(\cdot; \theta)$ and $\sigma(\cdot; \theta)$ are called the *drift* and the *diffusion coefficient* respectively and are allowed to depend on some parameters $\theta \in \Theta$. They are presumed to satisfy the regularity conditions (locally Lipschitz, with a linear growth bound) that guarantee the existence of a weakly unique, global solution of (4.1), see ch.4 of Kloeden and Platen (1992). In this paper we will consider only one-dimensional diffusions, although multivariate extensions are possible.

For sufficiently small time increment dt and under certain regularity conditions (see Kloeden and Platen, 1992), $V_{t+dt} - V_t$ is *approximately* Gaussian with mean and variance given by the so-called *Euler* (or Euler-Maruyama) approximation:

$$V_{t+dt} \approx V_t + b(V_t; \theta)dt + \sigma(V_t; \theta)\sqrt{dt}\mathcal{N}(0, 1),$$

though higher order approximations are also available. The *exact* dynamics of the diffusion process are governed by its transition density:

$$p_t(v, w; \theta) = \mathbb{P}[V_t \in dw \mid V_0 = v; \theta] / dw, \quad t > 0, w, v \in \mathbf{R}. \quad (4.2)$$

We shall assume that the process is observed without error at a given collection of time instances,

$$\mathbf{v} = \{V_{t_0}, V_{t_1}, \dots, V_{t_n}\}, \quad 0 = t_0 < t_1 < \dots < t_n;$$

this justifies the notion of a *discretely observed diffusion process*. The time increments between consecutive observations will be denoted $\Delta t_i = t_i - t_{i-1}$ for $1 \leq i \leq n$.

The log-likelihood of the data set $\mathbf{v}$ is:

$$\ell(\theta \mid \mathbf{v}) = \sum_{i=1}^n \ell_i(\theta), \quad \ell_i(\theta) := \log p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta).$$

Unfortunately, in all but a few special cases the transition density of the diffusion process and thus its likelihood are not analytically available. Therefore, it is already well documented that deriving maximum likelihood estimates (MLEs) for discretely observed diffusion processes is a very challenging problem. Nonetheless, theoretical properties of such MLEs are now well known in particular under ergodicity assumptions, see for example Kessler (1997); Gobet (2002).

Inference for discretely observed diffusions has been pursued in three main directions. One direction considers estimators alternative to the MLE. Established methods within this paradigm include techniques based on estimating functions (Bibby et al., 2002), indirect inference (Gourieroux et al., 1993) and efficient method of moments (Gallant and Long, 1997). Another direction involves numerical approximations to the unknown likelihood function. Aït-Sahalia (2002) advocated the use of closed-form analytic approximations to the unknown transition density; see Aït-Sahalia (2005) for multidimensional extensions. An alternative strategy has been to estimate an approximation to the likelihood using Monte Carlo (MC) methods. The approximation is given by Euler-type discretisation schemes, and the estimate is obtained using importance sampling. The strategy was put forward by Pedersen (1995a); Santa-Clara (1995) and was considerably refined by Durham and Gallant (2002). The third direction employs Bayesian imputation methods. The idea is to augment the observed data with values at additional time points so that a satisfactory complete-data likelihood approximation can be written down and to use the Gibbs sampler or alternative Markov chain Monte Carlo (MCMC) schemes; see Roberts and Stramer (2001); Elerian et al. (2001); Eraker (2001). An excellent review of several methods of inference for discretely observed diffusions is given in Sørensen (2004).

The approach introduced in this paper follows a different direction, which exploits recent advances in simulation methodology for diffusions. Exact simulation of diffusion sample paths has become feasible since the introduction of the Exact

Algorithm (EA) in Beskos et al. (2005c) [Article 3]. The algorithm is reviewed in Section 4.2 and relies on a technique called *retrospective sampling* developed originally in Papaspiliopoulos and Roberts (2004). To date there are two versions of the algorithm, EA1 which can be applied to a rather limited class of diffusion processes, which we call $\mathcal{D}_1$, and EA2 which is applicable to the much more general $\mathcal{D}_2$ class; all definitions are given in Section 4.2. The greater applicability of EA2 over EA1 comes at the cost of higher mathematical sophistication in its derivation, since certain results and techniques from stochastic analysis are required. However, its computer implementation is similar to that of EA1.

In this paper we show how to use EA in order to produce a variety of methods that can be used for maximum likelihood and Bayesian inference. We first discuss three unbiased MC estimators of the transition density (4.2) for a fixed value of θ : the bridge method (Section 4.4) first proposed in Beskos et al. (2005c) [Article 3], the acceptance method (Section 4.5) and the Poisson estimator (Section 4.6) first proposed in Wagner (1988b). The last two estimators are evolved in Sections 4.5.1 and 4.6 respectively to yield unbiased estimators of the transition density *simultaneously* for all $\theta \in \Theta$. Thus, the simultaneous estimators can readily be used in conjunction with numerical optimisation routines to estimate the MLE and other features of the likelihood surface.

We proceed by introducing a Monte Carlo EM (MCEM) algorithm in Section 4.8. The construction of the algorithm crucially depends on whether there are unknown parameters in the diffusion coefficient σ . The simpler case where only drift parameters are to be estimated is treated in Section 4.8.1, whereas the general case necessitates the path transformations of Roberts and Stramer (2001) and it is handled in Section 4.8.2.

Section 4.9 presents a MCMC algorithm which samples from the joint posterior distribution of the parameters and of appropriately chosen latent variables. Unlike currently favoured methods, our algorithm is not based on imputation of diffusion

paths but instead on what we call a *hierarchical simulation model*. In that way, our MCMC method circumvents computing the likelihood function.

Therefore, all our methodology is simulation-based, but it has advantages over existing methods of this type for two reasons.

1. The methods are *exact* in the sense that no discretisation error exists, and the MC estimation provides the only source of error in our calculations. Specifically, as the number of MC samples increases, the estimated MLE converge to the true MLE, and as the number of iterations in our MCMC algorithm increases, the samples converge to the true posterior distribution of the parameters.
2. Our methods are computationally efficient. Whilst approximate methods require rather fine discretisations (and consequently a number of imputed values which greatly exceeds the observed data size) to guarantee sufficient accuracy, our methodology suffers from no such restrictions.

A limitation of the methods introduced here is that their applicability is generally attached to that of EA. However, ongoing advances on EA itself (Beskos et al., 2005a) [Article 6] will weaken further the required regularity conditions so that a much larger class of diffusions than $\mathcal{D}_2$ can be effectively simulated. It is expected that these enhanced simulation algorithms will be of immediate use to the methods presented in this paper.

Our methods are illustrated on three different diffusion models. The first one is the periodic drift model, which belongs to $\mathcal{D}_1$ and, although it is interesting enough in its own right since its transition density is unavailable, it is used primarily for exposition. However, we also consider two more substantial and well-known applications, the logistic diffusion model for population growth, and the Cox-Ingersoll-Ross (CIR) model for interest rates. The former belongs to the $\mathcal{D}_2$ class, whereas the latter is a diffusion process outside the $\mathcal{D}_2$ class, and it is used to

illustrate how our exact methods can be extended for processes for which EA2 is not applicable. Moreover, since we can calculate analytically the likelihood for this model, we have a benchmark to test the success of our approach. We fit the CIR model to a well studied data set, which contains Eurodollar rates (recorded every ten days) between 1973 and 1995, to allow for comparisons with existing methods.

All the algorithms presented in this paper are coded in C and have been executed on a Pentium IV 2.6 GHz processor. We note that our methods are not computationally demanding according to modern statistical computing standards, and in the examples we have considered the computing times (reported explicitly in the following sections) were in the magnitude of seconds, or at worst minutes.

The structure of the paper is as follows. Section 4.2 reviews EA. Section 4.3 sets up the context of transition density estimation, Sections 4.4-4.6 present the three different estimators and Section 4.7 compares them theoretically and empirically. Section 4.8 introduces the MCEM algorithm and Section 4.9 the MCMC algorithm. We finish with some general conclusions and directions for further research in Section 4.10. Background material and proofs are collected in a brief Appendix.

4.2 Retrospective exact sampling of diffusions

Understanding the statistical methodology to be presented in later sections presupposes the introduction of the simulation techniques central to our approaches. In particular, we need to understand the form of the output provided by EA. The main references for the material of this section are Beskos et al. (2005c) [Article 3] and Beskos and Roberts (2005) [Article 2]. For the purposes of this paper it suffices to analyse EA for simulating diffusion paths conditional on their ending point (also known as diffusion bridges). In particular we will show how to simulate a diffusion path starting from $V_0 = v$ and ending at $V_t = w$, for any $t > 0$

and $v, w \in \mathbf{R}$. Simulation of unconditioned paths follows in the same lines and is sketched in Section 4.2.3. When necessary we characterise EA as *conditional* or *unconditional* to emphasise the type of diffusion simulation it is used for.

EA performs rejection sampling by proposing paths from processes we can simulate and accepting them according to appropriate probability density ratios. The novelty lies in the fact that the proposed paths are unveiled only at finite (but random) time instances and the decision whether to accept or not the path can be easily taken.

It is essential that we first transform the diffusion process (4.1) into an SDE of unit diffusion coefficient by applying the 1-1 transformation $V_s \mapsto \eta(V_s; \theta) =: X_s$, where

$$\eta(u; \theta) = \int^u \frac{1}{\sigma(z; \theta)} dz, \quad (4.3)$$

is any anti-derivative of $\sigma^{-1}(\cdot; \theta)$. Assuming that $\sigma(\cdot; \theta)$ is continuously differentiable, we apply Itô's rule to find that the SDE of the transformed process writes as:

$$dX_s = \alpha(X_s; \theta) ds + dW_s, \quad X_0 = x = \eta(V_0; \theta), \quad s \in [0, t], \quad (4.4)$$

where

$$\alpha(u; \theta) = \frac{b\{\eta^{-1}(u; \theta); \theta\}}{\sigma\{\eta^{-1}(u; \theta); \theta\}} - \sigma'\{\eta^{-1}(u; \theta); \theta\} / 2, \quad u \in \mathbf{R};$$

η^{-1} denotes the inverse transformation and σ' denotes the derivative w.r.t. the space variable. In the sequel we will make the following standard assumptions for any $\theta \in \Theta$:

1. $\alpha(\cdot; \theta)$ is continuously differentiable.
2. $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded below.
3. Girsanov's formula for X given in (4.32) is a martingale w.r.t. the Wiener measure.

We define

$$A(u; \theta) := \int^u \alpha(z; \theta) dz,$$

to be any anti-derivative of α . The transition density of X is defined as:

$$\tilde{p}_t(x, y; \theta) = \mathbf{P}[X_t \in dy | X_0 = x; \theta] / dy, \quad t > 0, x, y \in \mathbf{R}. \quad (4.5)$$

Before proceeding, we require the following notation. Let $C \equiv C([0, t], \mathbf{R})$ be the set of continuous mappings from $[0, t]$ to $\mathbf{R}$, $\mathcal{C}$ the corresponding cylinder σ -algebra and $\omega = \{\omega_s; s \in [0, t]\}$ a typical element of C . Let $\mathbb{Q}_\theta^{(t, x, y)}$ denote the distribution of the process X conditioned to start at $X_0 = x$ and finish at $X_t = y$, for some fixed x, y , and $\mathbb{W}^{(t, x, y)}$ be the probability measure for the corresponding Brownian bridge (BB). The notation highlights the dependence of the measure $\mathbb{Q}_\theta^{(t, x, y)}$ on θ .

The objective is to construct a rejection sampling algorithm to draw from $\mathbb{Q}_\theta^{(t, x, y)}$. The following lemma proved in Appendix is central to the methodology.

Lemma 4.1 *Under conditions 1-3 above, $\mathbb{Q}_\theta^{(t, x, y)}$ is absolutely continuous w.r.t. $\mathbb{W}^{(t, x, y)}$ with density:*

$$\frac{d\mathbb{Q}_\theta^{(t, x, y)}}{d\mathbb{W}^{(t, x, y)}}(\omega) = \frac{\mathcal{N}_{x, t}(y)}{\tilde{p}_t(x, y; \theta)} \exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\}.$$

Let

$$\begin{aligned} l(\theta) &\leq \inf_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta) / 2\}, \\ r(\omega; \theta) &\geq \sup_{s \in [0, t]} \{(\alpha^2 + \alpha')(\omega_s; \theta) / 2 - l(\theta)\}, \quad \omega \in C. \end{aligned}$$

In general r is a positive random variable. However in special cases it can be chosen independently of ω , for example if $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above. We now define the non-negative function $0 \leq \phi \leq 1$ which transforms a given path ω

as follows:

$$\phi(\omega_s; \theta) = \frac{1}{r(\omega; \theta)} \left\{ \frac{(\alpha^2 + \alpha')(\omega_s; \theta)}{2} - l(\theta) \right\}, \quad s \in [0, t]. \quad (4.6)$$

It is now clear that:

$$\frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(\omega) \propto \exp \left\{ -r(\omega; \theta) \int_0^t \phi(\omega_s; \theta) ds \right\} \leq 1, \quad \mathbb{W}^{(t,x,y)} - \text{a.s.} \quad (4.7)$$

Thus we have managed in (4.7) to bound the density ratio. The key to proceed to a feasible rejection sampler is to recognise (4.7) as a specific Poisson process probability.

Theorem 4.1 *Let $\omega \in C$, Φ be a homogeneous Poisson process of intensity $r(\omega; \theta)$ on $[0, t] \times [0, 1]$ and N be the number of points of Φ below the graph $s \mapsto \phi(\omega_s; \theta)$. Then*

$$\mathbb{P}[N = 0 | \omega] = \exp \left\{ -r(\omega; \theta) \int_0^t \phi(\omega_s; \theta) ds \right\}.$$

Theorem 4.1 suggests rejection sampling by means of an auxiliary Poisson process:

1. Simulate a sample path $\omega \sim \mathbb{W}^{(t,x,y)}$.
2. Calculate $r(\omega; \theta)$; generate a marked Poisson process $\Phi = \{\Psi, \Upsilon\}$, with points $\Psi = \{\psi_1, \dots, \psi_\kappa\}$ uniformly distributed on $[0, t]$ and corresponding marks $\Upsilon = \{v_1, \dots, v_\kappa\}$ uniformly distributed on $[0, 1]$, where $\kappa \sim \text{Po}(r(\omega; \theta)t)$.
3. Compute the acceptance indicator:

$$I := \prod_{j=1}^{\kappa} \mathbb{I}[\phi(\omega_{\psi_j}; \theta) < v_j]. \quad (4.8)$$

4. If $I = 1$, i.e. $\{N = 0\}$ has occurred, then accept ω , otherwise return to (1).

Unfortunately, this “algorithm” is impossible to implement since it requires the simulation of complete BBs on $[0, t]$. However, notice that it might be possible

to determine I based only on partial information about the proposed path. For instance, when r is only a function of θ (see Section 4.2.1) we can actually reverse the order in which Steps 1 and 2 are carried out. Specifically, we would first simulate Φ , and afterwards, *retrospectively*, we would realise ω at the time instances determined by Ψ , since this is sufficient for determining I . The technique of exchanging the order of simulation to implement in finite time simulation of infinite dimensional random variables has been termed *retrospective sampling* in Papaspiliopoulos and Roberts (2004).

The general framework under which EA operates assumes that it is possible to write:

$$r(\omega; \theta) = r(k(\omega); \theta),$$

where $k(\omega)$ has the following properties:

1. $k(\omega)$ is finite dimensional.
2. The law of $k(\omega)$ can be simulated (under $\mathbb{W}^{(t,x,y)}$).
3. The finite-dimensional distributions of $\mathbb{W}^{(t,x,y)}$ given $k(\omega)$ can be simulated.

Under these conditions, the following retrospective implementation of the rejection sampler can be carried out in finite time.

Exact Algorithm (EA)

1. **Simulate** $k(\omega)$.
 2. **Generate a realisation of the marked Poisson process** $\Phi = \{\Psi, \Upsilon\}$ of rate $r(k(\omega); \theta)$.
 3. **Simulate the skeleton** $\{\omega_{\psi_1}, \dots, \omega_{\psi_\kappa}\}$, conditionally upon $k(\omega)$.
 4. **Compute the acceptance indicator** I .
 5. **If** $I = 1$ **accept the proposed skeleton, and return** $k(\omega)$ **and**
 $S(\omega) := \{(0, x), (\psi_1, \omega_{\psi_1}), \dots, (\psi_\kappa, \omega_{\psi_\kappa}), (t, y)\}$; **else, return to (1)**.
-

$S(\omega)$ is an *exact* draw from a finite-dimensional distribution of $\mathbb{Q}_\theta^{(t,x,y)}$, which can be filled in at any required times afterwards using the Brownian bridge dynamics (see Section 4.2.3). The technical difficulty of finding or simulating $k(\omega)$, and consequently simulating the process at some time points given $k(\omega)$, imposes some restrictions on the applicability of EA. We now describe the two cases where the algorithm can be easily applied.

4.2.1 Exact Algorithm 1 (EA1)

Implementation of EA is straightforward when r does not depend on ω . This will be true within the following diffusion class.

Definition 4.1 *We say that a diffusion process V with SDE (4.1) belongs to $\mathcal{D}_1$, and write $V \in \mathcal{D}_1$, if the drift of the transformed process $X_s = \eta(V_s; \theta)$, $s \in [0, t]$, satisfies conditions 1–3 on pg.177, and $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above.*

In this case

$$r(\omega; \theta) \equiv r(\theta) = \sup_{u \in \mathbf{R}} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\} .$$

Within $\mathcal{D}_1$, Step 1 of EA is unnecessary, since no information about the path is *a priori* needed for determining the Poisson rate. Moreover, Step 3 entails simulation from a finite-dimensional distribution of BB (see Appendix A).

Example 1 (Periodic drift) Consider the following SDE:

$$dX_s = \sin(X_s - \theta)ds + dW_s .$$

Though apparently simple, this SDE cannot be solved analytically. However, EA1 can be applied since $X \in \mathcal{D}_1$ with $l(\theta) = -1/2$ and $r(\theta) = 9/8$. Our proposed methods will be tested on the **SINE** data set simulated from X with the unconditional version of EA1 (Beskos and Roberts, 2005) [Article 2] under the specifications $n = 1000$, $\Delta t_i = 1$, $X_0 = 0$, $\theta = \pi$; see also Fig.4.1. When θ is to be

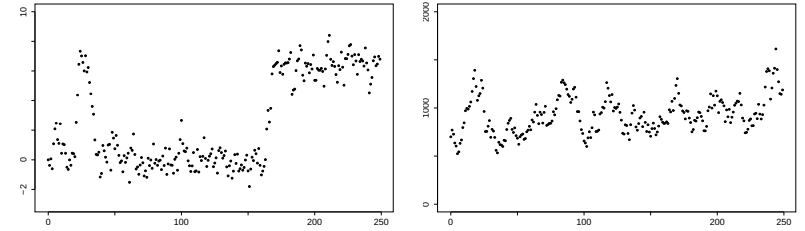


Figure 4.1: First 250 values of the **SINE** (left - using EA1) and the **LOG – GROWTH** (right - using EA2) data sets defined in Example 1, pg. 181, and Example 2, pg. 183, respectively.

estimated, we take $\Theta = [0, 2\pi]$ for identifiability reasons.

4.2.2 Exact Algorithm 2 (EA2)

EA2 applies to the wider class of diffusion processes defined below.

Definition 4.2 *We say that a diffusion process V with SDE (4.1) belongs to $\mathcal{D}_2$, and write $V \in \mathcal{D}_2$, if the drift of the transformed process $X_s = \eta(V_s; \theta)$, $s \in [0, t]$, satisfies conditions 1–3 on pg.177, and*

$$\limsup_{u \rightarrow \infty} (\alpha^2 + \alpha')(u; \theta) < \infty , \quad (4.9)$$

or $\limsup_{u \rightarrow -\infty} (\alpha^2 + \alpha')(u; \theta) < \infty$.

Due to symmetry, we will study only the case (4.9). We define the following elements of the path space $C = C([0, t], \mathbf{R})$:

$$m = \inf\{\omega_s; s \in [0, t]\}, \quad \tau = \sup\{s \in [0, t] : \omega_s = m\} ,$$

thus m is the minimum value of a path ω , and τ is the time that the minimum is attained. Within $\mathcal{D}_2$, $k(\omega) = m(\omega)$ and,

$$r(\omega; \theta) \equiv r(m; \theta) = \sup_{u \in [m, \infty]} \{(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)\} < \infty, \quad \omega \in C.$$

As shown in Beskos et al. (2005c) [Article 3] and summarised in Appendix A, m satisfies the three requirements in pg.180. Simulation of (τ, m) can be done using simple transformations of elementary random elements. It is known (Asmussen et al., 1995) that BB conditionally on (τ, m) , can be derived in terms of two independent *Bessel Bridges*, each operating on either side of (τ, m) . The Bessel bridge is defined as BB constrained to be positive and its simulation can be carried out by means of independent BBs; see Appendix A for details.

Example 2 (The logistic-growth model) A popular model for describing the dynamics of a population which grows at a geometric rate in an environment with limited feeding resources is given by the following SDE:

$$dV_s = R V_s (1 - V_s/C) ds + \beta V_s dW_s, \quad \theta = (R, C, \beta).$$

R is the growth rate per individual, C the maximal population supported by the resources of the environment and β a noise parameter; for details see ch.6 of Goel and Richter-Dyn (1974). Related models have been investigated in the context of financial economics, see e.g. Gouriéroux and Jasiak (2003). The transition density of V is not known analytically. The transformed process $X_s = -\log(V_s)/\beta$ solves the SDE:

$$dX_s = \left\{ \frac{\beta}{2} - \frac{R}{\beta} + \frac{R}{\beta C} \exp(-\beta X_s) \right\} ds + dW_s.$$

It can be verified that $V \in \mathcal{D}_2$ and that EA2 is applicable with $l(\theta) = \beta^2/8 - R/2$ and $r(m; \theta) = \max\{(\alpha^2 + \alpha')(m; \theta)/2 - l(\theta), R^2/(2\beta^2)\}$. Our proposed methods will be tested on the **LOG – GROWTH** data set simulated from V by first simulating

from X using the unconditional EA2 and then transforming $X_s \mapsto V_s$ (Beskos et al., 2005c) [Article 3]. We took $n = 1000$, $\Delta t_i = 1$, $V_0 = 700$ and parameters $(R, C, \beta) = (0.1, 1000, 0.1)$; see Fig.4.1.

4.2.3 Unconditional EA and path reconstruction

EA can be applied in a similar fashion when the diffusion is conditioned only on its initial point $X_0 = x$. The main difference lies in that the final value of the proposed path is distributed according to the density $h(u) \propto \exp\{A(u; \theta) - (u - x)^2/(2t)\}$ (which is assumed to be integrable). Simulation from h can be done efficiently using an adaptive rejection sampling approach. Conditionally on this final point, the rest of the path is a BB, and EA proceeds as already described.

The output of EA is a skeleton $S(\omega)$ and possibly the collection $k(\omega)$ of variables related to the path. However, we can afterwards fill in this finite representation of an accepted path ω according to the dynamics of the proposed BB and without further reference to the target process. For EA1, the path reconstruction requires the simulation of the BBs that connect the currently unveiled instances of an accepted path ω . For EA2, the extra conditioning of the proposed BB on its minimum m implies that the filling in will be driven by two independent Bessel bridges; see Appendix A.

4.3 Unbiased transition density estimators – preliminaries

As already discussed in the introduction one of our primary aims is to provide unbiased MC estimators of the transition density (4.2). The first step is to express (4.2) in terms of the density (4.5) of the transformed process X . A change of

variables argument yields

$$p_t(v, w; \theta) = \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta) \times |\eta'(w; \theta)|, \quad \text{for all } t, v, w, \theta.$$

Thus, we will demonstrate how to estimate $\tilde{p}_t(x, y; \theta)$ for arbitrary x, y, θ , using the *unconditional* EA (bridge method) and the *conditional* EA (acceptance method, Poisson estimator). Automatically, our methods yield unbiased estimators of the likelihood $\exp\{\ell(\theta \mid \mathbf{v})\}$, for any fixed $\theta \in \Theta$, so coupled with grid or more elaborate stochastic search algorithms they can be used for locating the MLE. However, this *pointwise* exploration of the likelihood surface is expected to be computationally inefficient due to the introduction of independent MC error at each likelihood evaluation, and it is not guaranteed to provide consistent MC estimators of the MLE. We overcome these drawbacks by modifying the two methods based on the conditional EA to give estimates of the likelihood function *simultaneously* for all $\theta \in \Theta$. The simultaneous methods achieve estimation of the complete function

$$\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta), \quad \text{for any fixed data points } v, w, \quad (4.10)$$

using a single stream of random elements which are independent of θ . Therefore, numerically efficient optimisation algorithms can be used to estimate features of the likelihood surface, such as the MLE and level sets. The simultaneous acceptance method possesses appealing consistency properties, as we discuss in Section 4.5.1.

A fundamental result which the acceptance method and the Poisson estimator are based on is the following corollary to Lemma 4.1. This important result is also contained in Lemma 1 of Dacunha-Castelle and Florens-Zmirou (1986).

Corollary 4.1 *Let X be the diffusion process in (4.4) with transition density*

$\tilde{p}_t(x, y; \theta)$. Then:

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_{x,t}(y) e^{A(y;\theta) - A(x;\theta)} \mathbb{E}_{\mathbb{W}^{(t,x,y)}} \left[\exp \left\{ - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\} \right]. \quad (4.11)$$

4.4 The bridge method

The bridge method introduced in Beskos et al. (2005c) [Article 3] uses explicitly the output of the *unconditional* EA to produce unbiased estimates of $\tilde{p}_t(x, y; \theta)$ for any fixed x, y, θ . We enrich the transition density expression to allow for conditioning upon further random elements. Let $\tilde{p}_t(x, y \mid \cdot; \theta) = \mathbb{P}[X_t \in dy \mid X_0 = x, \cdot; \theta] / dy$ be the conditional density of X_t given the starting value and any other random elements.

Consider some $\gamma > 0$. Suppose that $S(\omega)$, $k(\omega)$ are the output of the unconditional EA for simulating X started at $X_0 = x$ on the extended time interval $[0, t + \gamma]$. By conditional expectation properties, we have that:

$$\tilde{p}_t(x, y; \theta) = \mathbb{E}_{S(\omega), k(\omega)} [\tilde{p}_t(x, y \mid S(\omega), k(\omega); \theta)]. \quad (4.12)$$

The joint distribution of $S(\omega)$, $k(\omega)$ is intractable but can be easily simulated using EA on $[0, t + \gamma]$. Recall that the value of the path at times other than those specified by $S(\omega)$ can be obtained according to BB dynamics (in EA1) or Bessel bridge dynamics (in EA2). As a result, the conditional distribution of X_t given $S(\omega)$, $k(\omega)$, and thus the density on the right side of (4.12), can be easily identified. By construction $S(\omega)$ contains at least one time point on either side of t , since $(0, x), (t + \gamma, \omega_{t+\gamma}) \in S(\omega)$. Let t_- and t_+ denote the adjacent time points, $t_- < t < t_+$. By the Markov property $\tilde{p}_t(x, y \mid S(\omega), k(\omega); \theta)$ depends only on the points (t_-, ω_{t_-}) , (t_+, ω_{t_+}) of $S(\omega)$ and it will be equal either to a Brownian or a Bessel bridge density, both of which are analytically available and can be

computed explicitly.

Thus, the main attraction of the method is that the density of X_t given the simulated variables (the skeleton) is explicitly known, unlike for example the MC method by Pedersen (1995a), where it is approximated.

4.5 The acceptance method (AM)

In this section we show how to use a simple identity related to the *conditional* EA to derive an unbiased estimator of the transition density $\tilde{p}_t(x, y; \theta)$. Let $a(x, y, \theta)$ be the acceptance probability of EA for simulating from $\mathbb{Q}_\theta^{(t, x, y)}$. Directly, (4.7) implies that

$$a(x, y, \theta) = \mathbb{E}_{\mathbb{W}^{(t, x, y)}} \left[\exp \left\{ -r(\omega; \theta) \int_0^t \phi(\omega_s; \theta) ds \right\} \right]. \quad (4.13)$$

Substituting this expression in (4.11) and rearranging terms according to (4.6), yields the following identity which relates the acceptance probability of EA to the transition density of the diffusion:

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_{x, t}(y) \exp \{ A(y; \theta) - A(x; \theta) - l(\theta)t \} a(x, y, \theta). \quad (4.14)$$

Recall the definition of the acceptance indicator I in (4.8), which we will now rewrite as $I(x, y, \theta, \Phi, \omega)$ to emphasise the elements which determine its value: the Poisson process Φ , and the proposed bridge $\omega \sim \mathbb{W}^{(t, x, y)}$, unveiled at the times determined by Φ . By definition,

$$a(x, y, \theta) = \mathbb{E} [I(x, y, \theta, \Phi, \omega)], \quad (4.15)$$

where the expectation is taken with respect to the joint distribution of Φ and ω . Thus, a simple unbiased estimator of $a(x, y, \theta)$ for any fixed θ, x, y , can be

obtained by recording the number of times EA accepts a proposed skeleton in, K say, number of trials.

Estimation of the transition density through (4.14) and (4.15) is referred to as the acceptance method (AM). An essential feature of the resulting estimator is that it is a.s. bounded, since $I \leq 1$. As a result, all the moments of the estimator are finite.

4.5.1 The simultaneous acceptance method (SAM)

In this section we upgrade AM to yield an estimator of the complete map in (4.10). The method is now referred to as the simultaneous AM (SAM). We emphasise that (4.13)–(4.15) are applied for $x = x(\theta) = \eta(v; \theta)$ and $y = y(\theta) = \eta(w; \theta)$. The simultaneous estimator is obtained by expressing $a(x, y, \theta)$ as an expectation of a function of θ, x, y , and random variables none of which depends on θ, x and y . The method is presented for $V \in \mathcal{D}_1$, since the derivation of SAM for $V \in \mathcal{D}_2$ is technically much harder and it is contained in Beskos et al. (2005b) [Article 5], along with more theoretical results about the methods.

In AM, the Poisson process Φ and the proposed path ω both depend on θ , since the former is of rate $r(\theta)$ and the latter is a BB from $X_0 = x$ to $X_t = y$. As a result $a(x, y, \theta)$ can be estimated only by running a separate MC experiment for each $\theta \in \Theta$. Below we show how the thinning property of the Poisson process (see for example section 5.1 of Kingman, 1993) and the relocation-invariance property of BB (see Appendix A) can be exploited to decouple the dependence between Φ, ω and θ .

The relocation-invariance property of the BB suggests that we can rewrite the acceptance indicator in terms of a standard BB, $\omega \sim \mathbb{W}^{(t, 0, 0)}$, as follows:

$$I(x, y, \theta, \Phi, \omega) = \prod_{j=1}^{\kappa} \mathbb{I} \left[\phi \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) < v_j \right]. \quad (4.16)$$

Suppose we can find an $r_{\max} < \infty$, such that

$$r(\theta) \leq r_{\max}, \quad \text{for all } \theta \in \Theta.$$

Let $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be a marked Poisson process on $[0, t] \times [0, 1]$ with rate $r_{\max}$ and κ number of points, $\kappa \sim \text{Po}(r_{\max}t)$. The thinning property of the Poisson process implies that the process obtained by deleting each point of $\Phi_{\max}$ with probability $1 - r(\theta)/r_{\max}$ is a Poisson process with rate $r(\theta)$. Conditionally on κ , let $U = (U_1, \dots, U_\kappa)$ be a collection of iid $\text{Un}(0, 1)$ selection random variables. The thinning property implies that we can rewrite (4.16) in terms of $\Phi_{\max}$ as:

$$\begin{aligned} I(x, y, \theta, \Phi_{\max}, \omega, U) &= \\ &= \prod_{j=1}^{\kappa} \mathbb{I} \left[\mathbb{I} [U_j > r(\theta)/r_{\max}] \phi \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta \right) < v_j \right]; \end{aligned}$$

notice that only Poisson points for which $U_j < r(\theta)/r_{\max}$, determine the value of the indicator I .

Theorem 4.2 *Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and $\Phi_{\max} = \{\Psi_{\max}, \Upsilon_{\max}\}$ be an independent marked Poisson process of rate $r_{\max}$ with κ number of points. Then*

$$\begin{aligned} \mathbb{E} [I(x, y, \theta, \Phi_{\max}, \omega, U) \mid \Psi_{\max}, \omega] &= \\ (r_{\max})^{-\kappa} \prod_{j=1}^{\kappa} \left[r_{\max} + l(\theta) - \frac{1}{2}(\alpha^2 + \alpha') (\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y; \theta) \right] \end{aligned} \quad (4.17)$$

Notice that $a(x, y, \theta)$ is the expected value of (4.17) with respect to the joint distribution of $\Psi_{\max}$ and $\omega \sim \mathbb{W}^{(t,0,0)}$, thus (4.17) together with (4.14) determine an unbiased *simultaneous* estimator of $\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta)$, for any fixed data points v, w . An MC estimator of $a(x, y, \theta)$ is obtained by averaging over independent realisations of (4.17).

Example: Periodic drift. We applied SAM to the **SINE** data set choosing

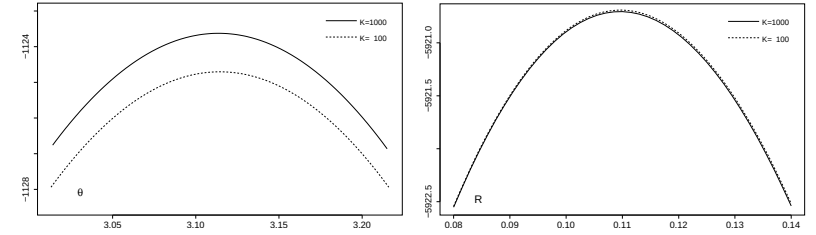


Figure 4.2: Simultaneous estimate of the log-likelihood for the **SINE** data set (left) and the profile log-likelihood (for the parameter R) for the **LOG-GROWTH** data set (right), using SAM with $K = 100$ (dashed) and $K = 1000$ (solid) MC samples.

$r_{\max} = 9/8$. For two different MC sample sizes, $K = 100$ and $K = 1000$, we used Brent's optimisation algorithm (which combines parabolic interpolation with the golden section search algorithm, see Section 10.2 of Press et al., 1992) for finding the maximum of the log-likelihood. The MLE was estimated as 3.116, when $K = 100$ (in 6 seconds), and 3.112 when $K = 1000$ (in 57 seconds). Using numerical differentiation, the standard error was estimated as 0.04 both when $K = 100$ and $K = 1000$. The left panel of Fig.4.2 shows the estimate of the log-likelihood function based on the two different MC sample sizes.

When the assumed model is in $\mathcal{D}_2$, it is mathematically much harder to derive a formula analogous to (4.17) which allows for simultaneous estimation of the likelihood function. In EA2 the minimum m and the time τ when the minimum occurs both depend on θ . Since the Poisson rate r is a function of m as well as θ , and ω is simulated conditionally on (τ, m) , the dependence of ω, Φ on θ is considerably harder than in EA1 to decouple and the method requires novel couplings and results from stochastic analysis. Such constructions are beyond the scope of this paper, and are presented in Beskos et al. (2005b) [Article 5], where several other issues related with SAM are also tackled. In particular it is proved that the simultaneous estimator for diffusions in $\mathcal{D}_1$ and $\mathcal{D}_2$ converges uniformly in θ to the likelihood function as the number of MC samples increases.

Under standard assumptions, this guarantees consistency of the MC estimate of the MLE.

Example: Logistic growth. We used SAM in conjunction with the downhill simplex optimisation method (section 10.4 of Press et al., 1992) to locate the MLE of the LOG – GROWTH data set, for MC sample sizes $K = 100$ and $K = 1000$. The estimates corresponding to (R, C, β) were $(0.1098, 1014.89, 0.10057)$ for $K = 100$ (in 2 minutes) and $(0.1097, 1014.78, 0.10057)$ for $K = 1000$ (in 20 minutes). Using numerical differentiation the curvature of the estimated log-likelihood at the estimated MLE was found:

$$\begin{pmatrix} 4567 & 0.072 & -9601 \\ & 0.011 & -0.05 \\ & & 198464 \end{pmatrix}$$

The corresponding standard errors are $(0.016, 9.5, 0.002)$. The plot at the right panel of Fig.4.2 shows the estimates of the profile log-likelihood of R for the two MC sample sizes.

4.6 The Poisson estimator

Corollary 4.1 relates the transition density of the diffusion to an expectation over the BB measure. Thus, any unbiased estimator of the expectation on the right-side of (4.11) corresponds to an unbiased estimator of $\tilde{p}_t(x, y; \theta)$. Generalizing, suppose that the expectation

$$\mathbb{E}_{\mathbb{P}^{(t,x,y)}} \left[\exp \left\{ - \int_0^t f(\omega_s) ds \right\} \right] < \infty, \quad (4.18)$$

is to be estimated for arbitrary continuous function f and diffusion bridge measure $\mathbb{P}^{(t,x,y)}$. For any $c \in \mathbf{R}$, $\lambda > 0$, and path ω , we write:

$$\begin{aligned} \exp \left\{ - \int_0^t f(\omega_s) ds \right\} &= e^{-ct} \sum_{j=0}^{\infty} \frac{1}{j!} \left\{ \lambda t \int_0^t \frac{c - f(\omega_s)}{\lambda t} ds \right\}^j = \\ &= e^{(\lambda-c)t} \mathbb{E} \left[\left\{ \int_0^t \frac{c - f(\omega_s)}{\lambda t} ds \right\}^{\kappa} \right], \end{aligned}$$

where the expectation is taken with respect to $\kappa \sim \text{Po}(\lambda t)$, conditionally on ω . Notice that if $\psi \sim \text{Un}(0, t)$, then $(c - f(\omega_\psi))/\lambda$ is an unbiased estimator of $\int_0^t (c - f(\omega_s))/(\lambda t) ds$. Let $\Psi = \{\psi_1, \dots, \psi_\kappa\}$ be a Poisson process of rate λ on $[0, t]$, and $\omega \sim \mathbb{P}^{(t,x,y)}$. Then, we obtain the following simple unbiased estimator of (4.18), which we call the *Poisson estimator*:

$$e^{(\lambda-c)t} \lambda^{-\kappa} \prod_{j=1}^{\kappa} [c - f(\omega_{\psi_j})]. \quad (4.19)$$

This estimator was introduced in the context of statistical physics by Wagner (1988a, 1989). It can be derived from first principles that its second moment is

$$e^{(\lambda-2c)t} \mathbb{E}_{\mathbb{P}^{(t,x,y)}} \left[\exp \left\{ \frac{1}{\lambda} \int_0^t [c - f(\omega_s)]^2 ds \right\} \right], \quad (4.20)$$

which is not guaranteed to be finite. Choice of c and λ with view of improving the efficiency of the algorithm is discussed in Section 4.7.

Taking $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, and $f = (\alpha^2 + \alpha')/2$, the Poisson estimator can be used with (4.11) to estimate $\tilde{p}_t(x, y; \theta)$. This transition density estimator will also be referred to as the Poisson estimator. A simultaneous estimation of the complete map $\theta \mapsto \tilde{p}_t(\eta(v; \theta), \eta(w; \theta); \theta)$ for any data points v, w , merely requires decoupling of ω in (4.19) from x and y , since Ψ clearly is independent of θ . Since in the current context $\mathbb{P}^{(t,x,y)} \equiv \mathbb{W}^{(t,x,y)}$, it suffices to exploit the relocation-invariance property

of BB and rewrite (4.19) as:

$$e^{(\lambda-c)t} \lambda^{-\kappa} \prod_{j=1}^{\kappa} \left[c - f \left(\omega_{\psi_j} + (1 - \psi_j/t)x + (\psi_j/t)y \right) \right], \quad \omega \sim \mathbb{W}^{(t,0,0)}. \quad (4.21)$$

This estimator will be referred to as the *simultaneous* Poisson estimator. For more general diffusion bridge measures, decoupling of ω from x, y can be cumbersome, since the intuitive relocation-invariance property does not hold in general. Such a case is treated in the following section.

4.6.1 Inference for the CIR model

We apply the Poisson estimator to infer about a diffusion process outside $\mathcal{D}_2$, the Cox-Ingersoll-Ross (CIR) diffusion model, which solves

$$dV_s = \rho(\mu - V_s) ds + \beta \sqrt{V_s} dW_s, \quad \theta = (\rho, \mu, \beta).$$

It is assumed that all parameters are positive and $2\rho\mu > \beta^2$, which guarantees that V does not hit zero (see pg.391 of Cox et al., 1985). We fit the CIR model to a well studied data set (used amongst others by Ait-Sahalia, 1996; Elerian et al., 2001; Roberts and Stramer, 2001) which contains daily Eurodollar rates between 1973 and 1995, to allow for comparisons with existing methods. We take a sub-sample of 550 values, corresponding to time intervals of 10 days, since the diffusion model seems more appropriate on that scale. We call this sub-sample the **EURODOLLAR** data set and plot it in the left panel of Fig.4.3.

In this case $\eta(u; \theta) = 2\sqrt{u}/\beta$, thus the transformation $X_s = \eta(V_s; \theta)$ solves the SDE

$$dX_s = \left(\left[\frac{2\rho\mu}{\beta^2} - \frac{1}{2} \right] \frac{1}{X_s} - \frac{\rho}{2} X_s \right) ds + dW_s.$$

Setting $\alpha(u; \theta) = (2\rho\mu/\beta^2 - 0.5)/u - \rho u/2$, it is easy to check that the CIR model is outside $\mathcal{D}_2$. X is a.s. positive and its distribution is absolutely continuous w.r.t.

that of the Bessel process. Let $\mathbb{Q}_\theta^{(t,x,y)}$ and $\mathbb{R}^{(t,x,y)}$ be the bridge measures of X and the Bessel process respectively, and $\tilde{p}_t(x, y; \theta)$ and $q_t(x, y)$ be the corresponding transition densities. Direct application of Lemma 4.1 yields an analogue of (4.11) for positive diffusions:

$$\tilde{p}_t(x, y; \theta) = \frac{x}{y} q_t(x, y) e^{A(y; \theta) - A(x; \theta)} \mathbb{E}_{\mathbb{R}^{(t,x,y)}} \left[\exp \left\{ - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\} \right]. \quad (4.22)$$

Thus, we can readily use the Poisson estimator to estimate the transition density of CIR for a fixed $\theta \in \Theta$. However, simultaneous estimation for all $\theta \in \Theta$ is non-trivial, since the decoupling of $\omega \sim \mathbb{R}^{(t,x,y)}$ from x, y is not straightforward. The appropriate construction is in fact a variation of the approach we have devised for SAM for diffusions in $\mathcal{D}_2$ and is contained in Beskos et al. (2005b) [Article 5]. This stems from the characterisation of the Bessel bridge as a BB conditioned to remain positive. As in EA2, we consider first the minimum m of the Bessel bridge and subsequently reconstruct the path given m . The distribution of m is the restriction on $(0, \infty)$ of the distribution of the minimum of the corresponding BB. Given m , the path can be reconstructed exactly as if it were BB with known minimum m .

We applied the simultaneous Poisson estimator to the **EURODOLLAR** data set taking, after some pilot tuning, $\lambda = c = 1$. At the right panel of Fig.4.3 we show the real and the estimated profile log-likelihood for β using the downhill simplex optimisation algorithm. Notice that even with 100 MC samples the estimated curve coincides with the true one, although it can be shown that the variance of the estimator is infinite in this case.

4.7 Comparison of transition density estimators

We have introduced three different methods which can be used to estimate the diffusion transition density. Whereas AM (and SAM) are devised solely for this

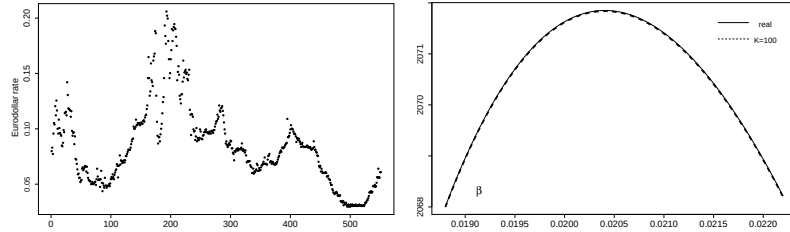


Figure 4.3: Left: The EURODOllAR data set. Right: The real (solid line) profile log-likelihood of β and its estimation (dashed line) using the simultaneous Poisson estimator with $\lambda = c = 1$ and $K = 100$ MC samples.

purpose, it is important to recognise that the other two methods have greater scope. The Poisson estimator can estimate expectations of general diffusion exponential functionals (4.18). The bridge method is based on (4.12) which is a form of *Rao-Blackwellisation* that can be used to estimate other conditional probabilities for diffusions, e.g. hitting probabilities.

In the context of transition density estimation, the weakness of AM and the bridge method over the Poisson estimator is that their applicability is determined by that of EA. On the contrary, the Poisson estimator can be used without assuming that the drift α is suitably bounded.

A significant advantage of AM over the competitors is that it is guaranteed to have finite polynomial moments. For the bridge method little is known since we have not as yet derived analytic expressions for its variance. Checking whether the variance of the Poisson estimator is finite is in general tedious. A notable exception is when the diffusion is in $\mathcal{D}_1$, where, since $f = (\alpha^2 + \alpha')/2$ is bounded, it is easy to see from (4.20) that the variance will be finite for any finite c, λ . Two examples outside $\mathcal{D}_1$ for which we know that this will not be the case are the CIR and the logistic growth models.

A feature of the bridge method which is likely to make it generally less efficient than the competitors is that the simulated paths ignore the final data point, since

the method is founded on the *unconditional* EA.

When the aim is to explore features of the likelihood surface, it is imperative that the estimators yield estimates of the transition density simultaneously for all $\theta \in \Theta$. The bridge method is the only for which we have been unable to achieve that.

Although derived from very different perspectives, SAM and the simultaneous Poisson estimator are related. Taking $\lambda = r_{\max}$, $c = \lambda + l(\theta)$, and contrasting (4.17) with (4.21), reveals that in $\mathcal{D}_1$, SAM is a special case of the simultaneous Poisson estimator. Moreover, there are certain optimality properties of this choice. For any c such that $c > r(\theta) + l(\theta)$, the second moment (4.20) is bounded above by $\exp\{(-2c + \lambda + (c - l(\theta))^2/\lambda)t\}$. This quantity is minimised by any pair (λ, c) such that $c = \lambda + l(\theta)$, where $\lambda \geq r(\theta)$. Requiring that the Poisson estimator yields estimates simultaneously for all $\theta \in \Theta$, the computationally most efficient bound on the variance is achieved by the choice $\lambda = r_{\max}$, $c = \lambda + l(\theta)$, under which the Poisson estimator and SAM coincide. Notice that in this choice, λ is the range and c the maximum of the functional $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}, \theta \in \Theta$. It is not obvious whether choosing $c > r(\theta) + l(\theta)$ is optimal. The connection between the two methods is less transparent outside $\mathcal{D}_1$. The rate of the Poisson process used in SAM for diffusions in $\mathcal{D}_2$, will depend on the minimum of the proposed path, and it is precisely because of this dependence that the estimator is a.s. bounded. The Poisson process and the proposed path in the Poisson estimator are inherently independent. Guided by our findings for $\mathcal{D}_1$, we propose to choose λ according to an estimate of the range of $(\alpha^2 + \alpha')(u; \theta)/2$ over all $u \in \mathbf{R}, \theta \in \Theta$, and take $c = \lambda + l(\theta)$. This choice has proved successful empirically.

A brief study of the performance of SAM and the simultaneous Poisson estimator in estimating the MLE of the LOG-GROWTH data set is summarised in Table 4.1. The parameters λ and c were chosen as suggested above. Taking into account its smaller computational cost, the simultaneous Poisson estimator is more

Parameter		Min.	1 st Quart.	Median	3 rd Quart.	Max.
R	SAM	0.1051	0.1083	0.1092	0.1100	0.1130
	SimP	0.1030	0.1072	0.1087	0.1099	0.1139
C	SAM	1010.13	1013.91	1014.89	1016.07	1021.07
	SimP	1008.88	1013.48	1014.93	1016.50	1022.80
β	SAM	0.10045	0.10053	0.10056	0.10058	0.10065
	SimP	0.10040	0.10050	0.10054	0.10058	0.10069

Table 4.1: Summary of 500 independent estimations of the MLE of the LOG – GROWTH data set. Each estimation is based on $K = 10$ MC samples. For each parameter, the first row corresponds to SAM and the second to the simultaneous Poisson estimator (SimP), with $\lambda = 1, c = l(\theta) + \lambda$. The average time for each experiment was around five times higher in SAM compared to the simultaneous Poisson estimator.

efficient in this example. A feature not depicted in the table is the sensitivity of the simultaneous Poisson estimator on the choice of c, λ . The method can produce unreliable estimates, especially with few MC samples, for certain choices of the tuning parameters. On the contrary, SAM is fully automatic.

4.8 A Monte Carlo EM (MCEM) approach

The EM algorithm (Dempster et al., 1977) is suitable for locating the MLE when the likelihood of the observed data is intractable but the joint likelihood of the observed and some appropriately defined missing data is of a simple form. At the expense of some additional effort the EM can also be used to obtain an estimate of the observed information matrix. To fix ideas, let $\mathbf{v}_{obs} = \{V_0 = v, V_t = w\}$ be the observed data, which consist of discrete-time observations from (4.1) (without loss of generality we assume two data points). It is assumed throughout this section that $V \in \mathcal{D}_2$. When the diffusion coefficient does not depend on unknown parameters, that is when $\sigma(\cdot; \theta) = \sigma(\cdot)$, we can treat the paths between consecutive observations as missing data and use the Girsanov's formula (4.32) to form the *complete likelihood*. This naturally leads to an EM algorithm whose implementation is described in the next section. On the other hand, when the diffusion

coefficient depends on θ , the missing paths should be appropriately transformed, as shown in Section 4.8.2.

4.8.1 MCEM for diffusions with known diffusion coefficient

We first transform the process $V_s \mapsto X_s = \eta(V_s)$, where η is defined in (4.3) and since by assumption $\sigma(\cdot; \theta) = \sigma(\cdot)$, η does not depend on θ . Thus, we define $\mathbf{x}_{obs} = \{X_0 = x, X_t = y\}$, $x = \eta(v)$, $y = \eta(w)$; $\mathbf{x}_{obs}$ contains discrete-time observations from the diffusion X with SDE (4.4). The observed log-likelihood is $\ell(\theta | \mathbf{x}_{obs}) = \log \tilde{p}_t(x, y; \theta)$. Let $\mathbf{x}_{mis}$ denote the missing path $\{X_s; s \in (0, t)\}$ and $\mathbf{x}_{com} = \{\mathbf{x}_{obs}, \mathbf{x}_{mis}\}$ denote the complete data, which is the continuous diffusion path starting from $X_0 = x$ and finishing at $X_t = y$. The complete log-likelihood is given by Girsanov's formula (4.32)

$$\ell(\theta | \mathbf{x}_{com}) = A(y; \theta) - A(x; \theta) - \frac{1}{2} \int_0^t (\alpha^2 + \alpha')(X_s; \theta) ds, \quad (4.23)$$

and

$$\mathbf{x}_{mis} | \mathbf{x}_{obs}; \theta \sim \mathbb{Q}_\theta^{(t, x, y)}.$$

The E-step of the EM algorithm requires analytic evaluation of

$$Q(\theta, \theta') = \mathbb{E}_{\mathbf{x}_{mis} | \mathbf{x}_{obs}; \theta'} [\ell(\theta | \mathbf{x}_{com})], \quad \theta \in \Theta.$$

However, notice that we can introduce the random variable $U \sim \text{Un}(0, t)$, which is independent of both $\mathbf{x}_{mis}$ and $\mathbf{x}_{obs}$, and re-write

$$Q(\theta, \theta') = A(y; \theta) - A(x; \theta) - \frac{t}{2} \cdot \mathbb{E}_{(\mathbf{x}_{mis}, U) | \mathbf{x}_{obs}; \theta'} [(\alpha^2 + \alpha')(X_U; \theta)]. \quad (4.24)$$

We are unable to perform analytically the expectation above. However, since we can simulate U , and X_u for any u using EA, we adopt an MC implementation of the

EM algorithm. The MCEM algorithm was introduced in Wei and Tanner (1990); convergence and implementation issues are tackled in Chan and Ledolter (1995); Sherman et al. (1999); Fort and Moulines (2003). It is well documented (see e.g. Fort and Moulines, 2003, and references therein) that the number of MC samples used to approximate the expectation should increase with the EM iterations. We have followed this approach in our examples. Several methods have been proposed for using the EM algorithm to estimate the observed information matrix; see for example Louis (1982); Jamshidian and Jennrich (2000); Meng and Rubin (1991). We have used the method suggested by Louis (1982), where we use MC estimations of the required expectations, as in Wei and Tanner (1990).

Example: Periodic drift. We applied MCEM to the **SINE** data set; the EM iterations are presented in the first column of Table 4.2. The M-step was implemented using Brent's optimisation method. The convergence is very rapid, essentially in one iteration, and the estimated MLE is in perfect agreement with the one obtained with SAM in Section 4.5.1. The standard error was estimated to be 0.04.

4.8.2 The general case: missing information and a path transformation

An important result from stochastic analysis, which has found numerous applications in statistical inference for diffusion processes (see for instance Barndorff-Nielsen and Shephard, 2002), is that a continuous-time diffusion path on an interval of time $[0, t]$ can be used to estimate perfectly the parameters involved in $\sigma(\cdot; \theta)$. A similar result does not hold for parameters in the drift $b(\cdot; \theta)$, where perfect estimation from continuous-time data typically holds asymptotically as $t \rightarrow \infty$. On the other hand, a finite number of discrete-time observations can contain only finite information about any of the parameters. The computational implication

is that we cannot construct an EM algorithm as in Section 4.8.1 where the paths between the observed data are treated as missing data. According to the EM terminology, this augmentation scheme leads to a fraction of missing information equal to 1. This problem was first encountered in an MCMC framework by Roberts and Stramer (2001) and Elerian (1999).

One remedy to this type of problems is to find a suitable invertible transformation of the missing data in order to reduce the augmented information. In particular, we want to transform the process so that the diffusion coefficient is independent of θ , exploiting the result that continuous-time paths contain only finite information about the drift parameters. Our approach is in the spirit of Roberts and Stramer (2001). We start by transforming the process, $V_s \mapsto X_s = \eta(V_s; \theta)$, where η is the transformation in (4.3). Therefore $X_0 = x(\theta) = \eta(v; \theta)$ and $X_t = y(\theta) = \eta(w; \theta)$. Notice that unlike Section 4.8.1, $\{X_s; s \in (0, t)\}$ is not directly observed, instead it is a function of the unknown parameters. The second level of path transformation is from $X_s \mapsto \dot{X}_s$ where:

$$\dot{X}_s := X_s - \left(1 - \frac{s}{t}\right)x(\theta) - \frac{s}{t}y(\theta), \quad s \in [0, t]. \quad (4.25)$$

$\dot{X}$ is a diffusion bridge starting from $\dot{X}_0 = 0$ and finishing at $\dot{X}_t = 0$. Its dynamics depend on θ and are typically intractable, nevertheless it is easy to simulate $\dot{X}_s$ at any time $s \in [0, t]$, conditionally upon $\mathbf{v}_{obs}$ and a specific value of θ , following the procedure:

1. Find $x = x(\theta)$, $y = y(\theta)$.
2. Simulate X_s from the bridge diffusion measure $\mathbb{Q}_\theta^{(t, x, y)}$ using EA.
3. Apply the transformation in (4.25).

The inverse transformation from $\dot{X}$ to X is:

$$g_\theta(\dot{X}_s) := \dot{X}_s + \left(1 - \frac{s}{t}\right)x(\theta) + \frac{s}{t}y(\theta), \quad s \in [0, t].$$

We propose an augmentation scheme with $\mathbf{v}_{mis} = \{\dot{X}_s; s \in [0, t]\}$ and $\mathbf{v}_{com} = \{\mathbf{v}_{obs}, \mathbf{v}_{mis}\}$. The following lemma, proved in Appendix B, gives the complete log-likelihood.

Lemma 4.2 *The log-likelihood of the complete data $\mathbf{v}_{com}$ is given by*

$$\begin{aligned} \ell(\theta | \mathbf{v}_{com}) &= \log |\eta'(w; \theta)| + \log \mathcal{N}_{x(\theta), t}(y(\theta)) \\ &+ A(y(\theta); \theta) - A(x(\theta); \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha') (g_\theta(\dot{X}_s); \theta) \, ds. \end{aligned}$$

Arguing as in Section 4.8.1, we have that

$$\begin{aligned} Q(\theta, \theta') &= \log |\eta'(w; \theta)| + \log \mathcal{N}_{x(\theta), t}(y(\theta)) \\ &+ A(y(\theta); \theta) - A(x(\theta); \theta) - \frac{t}{2} \mathbb{E}_{(\mathbf{v}_{mis}, U) | \mathbf{v}_{obs}; \theta'} [(\alpha^2 + \alpha') (g_\theta(\dot{X}_U); \theta)]. \end{aligned}$$

We can now apply MCEM in a similar fashion as in Section 4.8.1.

Example: Logistic growth. We used MCEM to estimate the MLE and the observed information for the **LOG – GROWTH** data set. The results in Table 4.2 indicate rapid convergence. We used the Fletcher-Reeves version of the conjugate gradient method for the M-step of the algorithm (see section 10.5 of Press et al., 1992). The standard errors are estimated using $K = 2000$ as (0.015, 9.5, 0.002).

4.9 Hierarchical simulation models and inference using MCMC

In this section we develop an MCMC algorithm for Bayesian inference for discretely observed diffusions. The stationary distribution of our MCMC is the *exact* posterior distribution of the parameters. However, it is not only the *exactness* which distinguishes our approach from competitive existing MCMC algorithms. Existing methods follow a path augmentation approach, as we did in the EM algorithm in

Iteration j	$\theta^{(j)}$	$R^{(j)}$	$C^{(j)}$	$\beta^{(j)}$
0	0.500	0.7000	1500.00	0.7000
1	3.119	0.1099	1015.32	0.1006
2	3.114	0.1097	1014.80	0.1006
3	3.115	0.1097	1014.83	0.1006
4	3.115	0.1097	1014.51	0.1006
5	3.114	0.1097	1014.66	0.1006
6	3.113	0.1097	1014.76	0.1006
7	3.113	0.1097	1014.67	0.1006
8	3.113	0.1097	1014.75	0.1006
9	3.113	0.1097	1014.71	0.1006
10	3.113	0.1095	1014.78	0.1006

Table 4.2: MCEM iterations for the **SINE** (first column) and the **LOG – GROWTH** (last three columns) data sets. The first five iterations were performed using 200 MC samples and the last five using 2000 samples. The execution times were 177 seconds for **SINE** and 18 minutes for **LOG – GROWTH**.

Section 4.8. The missing continuous paths are approximated by a fine discrete-time Markov chain, whose transition is assumed to follow an Euler-type approximation to the true diffusion transition. The joint distribution of the observed and missing data is given by an appropriate approximation of Girsanov’s formula. Then, a Gibbs sampler (or more general a component-wise updating MCMC) is used to sample from the approximate posterior distribution of the parameters and the missing paths. Often, for any two data points several points need to be imputed in between them to obtain a good approximation to the true posterior of θ . However the performance of basic MCMC schemes can severely deteriorate as the amount of imputation increases; see Roberts and Stramer (2001) for details.

The approach introduced here is not based on augmentation of paths. Instead, we construct a graphical model which involves the variables used in EA, and we show that the posterior distribution of the parameters is obtained as a marginal in this graphical model. Thus, we use an appropriate Metropolis-Hastings algorithm to sample from the joint posterior of all variables in the graph. One of the steps of the sampler involves running the *conditional* EA. The state-space of our MCMC typically has much smaller dimension than the competing augmentation methods

and as a result it can be computationally more efficient. However, comparison between alternative MCMC methods is not carried out in this paper, and will be reported elsewhere.

We describe in detail the MCMC algorithm when it is assumed that $V \in \mathcal{D}_1$. The derivation of the algorithm when $V \in \mathcal{D}_2$ is technically much harder, and it can be found in Beskos et al. (2006a). Nevertheless, we present simulation results for both cases. An essential ingredient of the algorithm is the following lemma which derives the density of the output of the conditional EA1.

Lemma 4.3 *Consider any two fixed points x and y . Let $\Phi = \{\Psi, \Upsilon\}$ be the marked Poisson process on $[0, t] \times [0, 1]$ with rate $r(\theta)$ and number of points $\kappa \sim \text{Po}(r(\theta)t)$, which is used in EA1 for simulating from $\mathbb{Q}_\theta^{(t,x,y)}$. Let $\omega \sim \mathbb{W}^{(t,0,0)}$, and I be the acceptance indicator (4.16) which decides if $\{\omega_s + (1-s/t)x + (s/t)y; s \in [0, t]\}$ is accepted as a path from $\mathbb{Q}_\theta^{(t,x,y)}$. Then, the conditional density of ω and Φ given $\{I = 1\}$, $\pi(\omega, \Phi \mid \theta, x, y, I = 1)$, is*

$$\frac{\exp\{t[1 - r(\theta)]\} r(\theta)^\kappa}{a(x, y, \theta)} \prod_{j=1}^{\kappa} \mathbb{I} \left[\phi \left(\omega_{\psi_j} + \left(1 - \frac{\psi_j}{t}\right)x + \frac{\psi_j}{t}y; \theta \right) < v_j \right], \quad (4.26)$$

w.r.t. the product measure $\mathbb{W}^{(t,0,0)} \times \Phi$, where Φ is the measure of a unit-rate Poisson process on $[0, t] \times [0, 1]$, and $a(x, y, \theta)$ is the acceptance probability of EA1.

To clarify the notation, in the remainder of the section (ω^*, Φ^*) represent the accepted random elements in EA1, with density $\pi(\omega^*, \Phi^* \mid \theta, x, y)$ given in (4.26); Φ^* , and ω^* at any collection of times, can be easily sampled using the conditional EA1.

Let $\pi(\theta)$ denote an appropriately specified prior density for θ . The aim is to sample from the posterior distribution of θ given the diffusion observations $\mathbf{v} = \{V_{t_0}, \dots, V_{t_n}\}$, say $\pi(\theta \mid \mathbf{v})$, where $V \in \mathcal{D}_1$. It is convenient to define

$$x_i = x_i(\theta) := \eta(V_{t_i}; \theta), \quad i = 0, \dots, n, \quad (4.27)$$

where η is the transformation in (4.3). For a given θ , the x_i 's are discrete-time samples from the SDE (4.4). For $1 \leq i \leq n$, let $\omega_i^* = \{\omega_{i,s}^*; s \in [t_{i-1}, t_i]\}$ and $\Phi_i^* = \{\Psi_i^*, \Upsilon_i^*\}$ be the accepted elements of the conditional EA1 applied on the time interval $[t_{i-1}, t_i]$; $S(\omega_i^*)$ will denote the corresponding output skeleton. We construct the following hierarchical model, defined through the densities:

$$\begin{aligned} \theta &\sim \pi(\theta) \\ V_{t_i} \mid V_{t_{i-1}}, \theta &\sim p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta) \\ (\omega_i^*, \Phi_i^*) \mid \theta, x_{i-1}(\theta), x_i(\theta) &\sim \pi(\omega_i^*, \Phi_i^* \mid \theta, x_{i-1}(\theta), x_i(\theta)). \end{aligned} \quad (4.28)$$

Notice that in this hierarchical model the observed data are in the middle of the hierarchy. We term (4.28) a *hierarchical simulation model*, to emphasise that it represents the order in which variables have to be simulated to ensure that the output of EA1, ω_i^* , is indeed from $\mathbb{Q}_\theta^{(\Delta t_i, x_{i-1}, x_i)}$. The posterior density of interest, $\pi(\theta \mid \mathbf{v})$, is a marginal of the joint posterior density of θ and the latent variables, $\pi(\theta, \{\omega_i^*, \Phi_i^*; 1 \leq i \leq n\} \mid \mathbf{v})$. We aim at sampling from this density via the Gibbs sampler. The conditional density of the latent variables is

$$\pi(\{\omega_i^*, \Phi_i^*; 1 \leq i \leq n\} \mid \theta, \mathbf{v}) = \prod_{i=1}^n \pi(\omega_i^*, \Phi_i^* \mid \theta, x_{i-1}(\theta), x_i(\theta)), \quad (4.29)$$

thus the pairs (ω_i^*, Φ_i^*) are conditionally independent with density given in (4.26). The key property of this hierarchical simulation model is that θ is independent of $\{(\omega_i^*, \Phi_i^*); 1 \leq i \leq n\}$ given the collection of skeletons $\{S(\omega_i^*); 1 \leq i \leq n\}$. Specifically, the conditional density of θ given the observed data and the latent variables is given in the following theorem.

Theorem 4.3 *θ is conditionally independent of $\{(\omega_i^*, \Phi_i^*); 1 \leq i \leq n\}$ given $\mathbf{v}$ and*

$\{S(\omega_i^*); 1 \leq i \leq n\}$, with density $\pi(\theta \mid \{S(\omega_i^*); 1 \leq i \leq n\}, \mathbf{v})$ proportional to:

$$\begin{aligned} & \pi(\theta) r(\theta)^{\sum \kappa_i^*} \exp \left\{ A(x_n(\theta); \theta) - A(x_0(\theta); \theta) - [l(\theta) + r(\theta)] \sum_{i=1}^n \Delta t_i \right\} \times \\ & \prod_{i=1}^n |\eta'(x_i(\theta); \theta)| \mathcal{N}_{x_{i-1}(\theta), \Delta t_i}(x_i(\theta)) \times \\ & \prod_{i=1}^n \left\{ \prod_{j=1}^{\kappa_i^*} \left[1 - \phi \left(\omega_{i, \psi_{i,j}^*}^* + (1 - \psi_{i,j}^* / \Delta t_i) x_{i-1}(\theta) + (\psi_{i,j}^* / \Delta t_i) x_i(\theta); \theta \right) \right] \right\}. \end{aligned} \quad (4.30)$$

Notice that (4.30) is given in a simple computable form, and the intractable terms $a(x_{i-1}, x_i, \theta)$ are not involved when conditioning upon $\{S(\omega_i^*); 1 \leq i \leq n\}$. Our proposed MCMC algorithm is a component-wise updating algorithm, simulating iteratively the skeletons $\{S(\omega_i^*); 1 \leq i \leq n\}$ and θ from their posterior conditional distributions. The skeletons $\{S(\omega_i^*); 1 \leq i \leq n\}$ are conditionally independent given θ and can be easily simulated using EA1. In some cases it might be possible to sample from (4.30) directly, otherwise a Metropolis-Hastings step could be used. When a random walk Metropolis is being used, it will generally make sense to scale the proposal variance to be proportional to n^{-1} .

Example: Periodic drift. We implemented the algorithm for the **SINE** data set. We used a uniform prior on $[0, 2\pi]$. It took 47 seconds to run the algorithm for 10,000 iterations, and some summaries are shown in the rightmost column of Fig.4.4. The posterior mean is estimated as 3.1127 and the posterior standard deviation as 0.04. We used a Metropolis step to update θ which had acceptance probability 0.49. The algorithm mixes very rapidly, and essentially the autocorrelation in the θ series is due to the fact that a Metropolis step is used rather than direct simulation from its conditional distribution. The dependence between θ and the latent variables is very weak.

When the assumed model is in $\mathcal{D}_2$, the construction of the MCMC algorithm is

more complicated. In particular, derivation of the joint density of $(\theta, \omega^*, \Phi^*)$ is challenging due to the more complex structure of EA2. This is done in Beskos et al. (2006a), where also other important issues related with the implementation of the MCMC are tackled. The algorithm necessitates recent non-centring reparametrisation methodology for hierarchical models as described in Roberts et al. (2004); Papaspiliopoulos et al. (2003).

Example: Logistic growth. Using the MCMC algorithm described in Beskos et al. (2006a), we obtained 50,000 samples from the posterior distribution of the parameters (R, C, β) for the **LOG – GROWTH** data set. The computing time was 20 minutes, and summary of the results is given in Fig.4.4. The posterior means were estimated as (0.1075, 1017.4, 0.1007), the posterior precision matrix as

$$\begin{pmatrix} 4589 & 0.065 & -9408 \\ & 0.001 & 0.116 \\ & & 196060 \end{pmatrix}$$

from which the posterior standard deviations read as (0.015, 31.13, 0.002).

4.10 Conclusions

In this paper we have introduced a variety of methods which can be used for likelihood-based inference for discretely observed diffusions. The methods rely on recent advances in the exact simulation of diffusions. The computational efficiency of the methods was illustrated in a collection of examples. However, an exhaustive simulation study which tests the relative performance of our methods and existing approaches, under various model specifications and parameter settings, is not given here. Such a detailed empirical investigation is currently taking place. However, some qualitative remarks are made below.

In general, the computing time required for our methods critically depends on

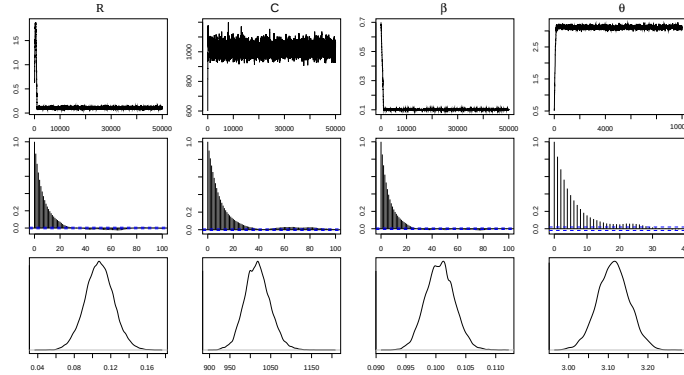


Figure 4.4: Summary of MCMC results: trace plots (top row), autocorrelation plots (middle) and posterior density estimates (bottom). Last column: 10,000 iterations of MCMC for the *SINE* data set (first 500 samples removed as burn-in for autocorrelation and density estimation). First three columns: 50,000 iterations of MCMC for the *LOG-GROWTH* data set (burn-in of 2,000 iterations).

the rate of the Poisson process. *Ceteris paribus*, this rate is a function of the time increment Δt between the observations. The performance (measured either in computing time or MC error) of our methods will be very strong for small Δt (high frequency data). However, the performance of the methods as presented here, deteriorates for sparser data sets (Δt large). Specifically, the problem stems from the following two characteristics of the conditional EA: i) the Poisson rate is linear in Δt (in EA2 it can increase even faster), ii) the acceptance rate typically decreases exponentially to 0 in Δt . The computing time required for AM and SAM increases due to i) and for MCEM and MCMC due to i) and ii). In AM and SAM, the variance will also increase with Δt , since MC averages would be heavily dominated by the terms (of very small probability) corresponding to Poisson configurations with very few points. Similarly, (4.20) suggests that the variance of the Poisson estimator increases exponentially with Δt . In MCMC, the computing cost can be transformed to be linear in Δt by augmenting any two observed data with additional points in between and applying the conditional EA on intervals of length

a fraction of Δt . Of course, the additional augmentation will affect the MCMC mixing.

A number of extensions of our methodology are possible. One important direction is the extension of our methods for diffusions outside the $\mathcal{D}_2$ class. As we have shown, the Poisson estimator can be readily used for likelihood inference for more general diffusion processes. On the other hand, current progress on EA itself, which attempts to remove the boundedness conditions on the drift (Beskos et al., 2005a) [Article 6] are expected to broaden the applicability of the estimation methods. It is worth noticing, that MCEM and AM remain unaltered irrespective of which version of EA is being used, thus they can readily accommodate extensions in EA. The other methods use explicitly the structure of EA and will have to be modified appropriately.

We have concentrated here on the case where the diffusion is observed without error at a finite set of times. However, data often occurs in different forms. For instance, data might be subject to observation error. All the methods for evaluating MLEs are difficult to extend to this case. However, the MCMC approach can be extended in a straightforward manner. An alternative interesting data form is when we observe a one-dimensional component of a higher dimensional diffusion, as for instance in continuous-time filtering models. We are currently working on a collection of such filtering problems, and we have found that our methods can be extended to this case, although there are significant additional implementation challenges in this approach.

Our methodology extends some way to time-inhomogeneous and multivariate diffusions. In carrying out these extensions, there are two steps in the arguments used in this paper that need to be generalised. Firstly, it is necessary to generalise the transformation (4.3) which eliminates the diffusion coefficient. This is straightforward under mild smoothness conditions on σ in the time-inhomogeneous extension. For multivariate extensions, however, the generalised version of (4.3)

involves the solution of an appropriate vector differential equation which is often intractable or insolvable (see e.g. Aït-Sahalia, 2005). This imposes restrictions on the class of multivariate diffusions to which the methods presented here can currently be applied. Secondly, we need to eliminate the stochastic integral from Girsanov's formula (4.32) to derive Lemma 4.1. Again this is fairly routine in the time-inhomogeneous case, while an extra condition is needed in the multivariate case, requiring that the multivariate drift be the gradient of a suitable potential function. This is a well-known condition in stochastic analysis and for ergodic, unit-diffusion coefficient diffusions corresponds to reversibility.

It is natural to ask whether the ideas of this paper extend to SDEs driven by Lévy processes. However, the Cameron-Martin-Girsanov formula, providing a closed form likelihood function, is critical to all our methodology. Unfortunately, there does not exist an analogous expression for the Radon-Nikodym derivative of an infinite activity Lévy driven SDE with respect to a tractable and easy to simulate from measure. Nevertheless, it is straightforward to extend our methods to incorporate SDEs with jumps according to a finite activity Lévy process (as for example considered in Roberts et al., 2004).

We hope that the collection of techniques described in this paper will have further applications. We are currently working on MC estimation of derivative hedge ratios in finance. The methodology builds on Rao-Blackwellisation techniques such as those devised for the bridge method.

Whilst we recognise that the mathematical details of this work are complex for those without a working knowledge of diffusion theory, we firmly believe that our methods have the potential to influence applied statistical work. This is because the algorithms we use are relatively simple and easy to code, the methods are not computationally demanding and are capable of handling long time-series. In addition, motivated by the desire to make our work as accessible as possible, we have started developing generic software for the implementation of some of our

methods, beginning with EA.

Acknowledgements

The authors would like to thank Mike Pitt and Neil Shephard for helpful suggestions. We are grateful to all referees for their comments which contributed to the improvement of the paper. The first author acknowledges financial support from the Greek State Scholarship's Foundation and the second author is funded by EPSRC grant GR/S61577/01.

APPENDIX

Appendix A: Background Material

1 . Brownian motion and Brownian bridge

A standard Brownian motion $\{W_s; s \in [0, t]\}$, $W_0 = 0$, can be simulated on a collection of time instances $0 = s_0 < s_1 < \dots < s_k \leq t$ in the following recursive way:

$$W_{s_j} = W_{s_{j-1}} + Y_j, \quad Y_j \sim \mathcal{N}(0, s_j - s_{j-1}), \quad 1 \leq j \leq k.$$

A standard BB with distribution $\mathbb{W}^{(t,0,0)}$ can be obtained as the transformation:

$$W_s - \frac{s}{t}W_t, \quad s \in [0, t].$$

An arbitrary BB with distribution $\mathbb{W}^{(t,x,y)}$ can be simulated via the relocation-invariance property of BB: if $\omega \sim \mathbb{W}^{(t,0,0)}$, then $\omega_s + (1 - s/t)x + (s/t)y$, is a path from $\mathbb{W}^{(t,x,y)}$.

2. Decomposition of a Brownian path at its minimum

The following theorem, stated in Beskos et al. (2005c) [Article 3], shows how to simulate a BB path together with its minimum m and the time τ when the minimum is attained.

Theorem 4.4 *A path $\omega \sim \mathbb{W}^{(t,x,y)}$ can be simulated in the following way:*

1. Simulate m, τ according to their joint density:

$$\pi(m, \tau) = \sqrt{2\pi t} \frac{(y-m)(x-m)}{\sqrt{\tau^3(t-\tau)^3}} \exp \left\{ -\frac{(x-m)^2}{2\tau} - \frac{(y-m)^2}{2(t-\tau)} + \frac{(y-x)^2}{2t} \right\}, \quad (4.31)$$

for $m \leq \min\{x, y\}$ and $\tau \in [0, t]$.

2. The rest of the path is a transformation of six independent standard BBs,

$b_j \sim \mathbb{W}^{(1,0,0)}$, $j = 1, \dots, 6$. Analytically, for $s \in [0, \tau]$,

$$\omega_s = m + \sqrt{\tau} \cdot \sqrt{\left(\frac{(x-m)(\tau-s)}{\tau^{3/2}} + b_{1,(1-s/\tau)} \right)^2 + b_{2,(1-s/\tau)}^2 + b_{3,(1-s/\tau)}^2},$$

and for $s \in [\tau, t]$,

$$\omega_s = m + \sqrt{t-\tau} \cdot \sqrt{\left(\frac{(y-m)(s-\tau)}{(t-\tau)^{3/2}} + b_{4,(s-\tau)/(t-\tau)} \right)^2 + b_{5,(s-\tau)/(t-\tau)}^2 + b_{6,(s-\tau)/(t-\tau)}^2}.$$

In Step 1, m is drawn from its marginal distribution, which is a transformed Rayleigh distribution, and conditionally on m , τ is drawn using the inverse CDF method.

3. Bessel process and Bessel bridge

The (3-dimensional) Bessel process/bridge can be defined as a scalar Brownian motion/bridge constrained to be positive; for main results see ch.11 of Revuz and Yor (1991). Recall that an arbitrary Bessel bridge measure is denoted by $\mathbb{R}^{(t,x,y)}$.

Simulation of $\omega \sim \mathbb{R}^{(t,0,y)}$ can be carried out by means of three independent BBs (see e.g. Bertoin and Pitman, 1994):

$$\omega_s = \sqrt{(ys/t + b_{1,s})^2 + b_{2,s}^2 + b_{3,s}^2}, \quad s \in [0, t], \quad b_j \sim \mathbb{W}^{(t,0,0)}.$$

Simulation of $\omega \sim \mathbb{R}^{(t,x,y)}$ for $x, y > 0$ is more involved and follows the lines of Theorem 4.4. We first simulate its minimum m , whose distribution is the restriction on $(0, \infty)$ of the distribution of the minimum of the corresponding BB. Given m , the path can be reconstructed exactly as if it were BB with known minimum m .

Appendix B: Proofs of main results

Proof of Lemma 4.1:

Girsanov's formula, which gives the density of the law of X w.r.t. the Wiener measure, is

$$\exp \left\{ \int_0^t \alpha(\omega_s; \theta) d\omega_s - \frac{1}{2} \int_0^t \alpha^2(\omega_s; \theta) ds \right\}. \quad (4.32)$$

Thus, using Itô's Lemma

$$\begin{aligned} \frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(\omega) &= \frac{\mathcal{N}_{x,t}(y)}{\tilde{p}_t(x, y; \theta)} \exp \left\{ \int_0^t \alpha(\omega_s; \theta) d\omega_s - \frac{1}{2} \int_0^t \alpha^2(\omega_s; \theta) ds \right\} \\ &= \frac{\mathcal{N}_{x,t}(y)}{\tilde{p}_t(x, y; \theta)} \exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\}. \end{aligned}$$

Proof of Lemma 4.2:

We will find the density $\pi(V_t, \dot{X} | \theta)$ of the law of $\mathbf{v}_{com}$ with respect to $\text{Leb} \times \mathbb{W}^{(t,0,0)}$, where Leb denotes the Lebesgue measure on $\mathbf{R}$. Clearly,

$$\pi(V_t, \dot{X} | \theta) = p_t(V_0, V_t; \theta) \times \pi(\dot{X} | V_t, \theta),$$

The law of the 1-1 transformation $g_\theta(\dot{X})$ given V_t is just $\mathbb{Q}_\theta^{(t,x,y)}$, where $x = \eta(V_0; \theta)$ and $y = \eta(V_t; \theta)$. Thus,

$$\pi(\dot{X} \mid V_t, \theta) = \frac{d\mathbb{Q}_\theta^{(t,x,y)}}{d\mathbb{W}^{(t,x,y)}}(g_\theta(\dot{X})) .$$

Use of Lemma 4.1 concludes the proof.

Proof of Lemma 4.3:

Notice that the Radon-Nikodym derivative of a homogeneous Poisson process with rate $r(\theta)$ on $[0, t] \times [0, 1]$ w.r.t. Φ is:

$$\exp\{-r(\theta)t\} \frac{r(\theta)^\kappa t^\kappa}{\kappa!} \times \left[\exp\{-t\} \frac{t^\kappa}{\kappa!} \right]^{-1} ,$$

which simplifies to $\exp\{t[1 - r(\theta)]\}r(\theta)^\kappa$.

Proof of Theorem 4.3:

The joint density of θ and the latent variables conditionally on $\mathbf{v}$ can be decomposed as

$$\pi(\theta, \{\omega_i^*, \Phi_i^*; 1 \leq i \leq n\} \mid \mathbf{v}) = \pi(\theta \mid \mathbf{v}) \times \pi(\{\omega_i^*, \Phi_i^*; 1 \leq i \leq n\} \mid \theta, \mathbf{v}) .$$

To simplify the formulae, we write x_i instead of $x_i(\theta)$. Clearly,

$$\pi(\theta \mid \mathbf{v}) \propto \pi(\theta) \prod_{i=1}^n \tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta) |\eta'(x_i; \theta)| .$$

Using (4.14), the acceptance probability of EA1 $a(x_{i-1}, x_i, \theta)$, can be expressed as:

$$\tilde{p}_{\Delta t_i}(x_{i-1}, x_i; \theta) \cdot \left[\mathcal{N}_{x_{i-1}, \Delta t_i}(x_i - x_{i-1}) \exp\{A(x_i; \theta) - A(x_{i-1}; \theta) - l(\theta)\Delta t_i\} \right]^{-1} .$$

Combining this expression with (4.29) and (4.26) we find that the joint density $\pi(\theta, \{\omega_i^*, \Phi_i^*; 1 \leq i \leq n\} \mid \mathbf{v})$ is proportional to:

$$\begin{aligned} & \pi(\theta) r(\theta)^{\sum \kappa_i^*} \exp \left\{ A(x_n; \theta) - A(x_0; \theta) - [l(\theta) + r(\theta)] \sum_{i=1}^n \Delta t_i \right\} \times \\ & \prod_{i=1}^n |\eta'(x_i; \theta)| \mathcal{N}_{x_{i-1}, \Delta t_i}(x_i) \times \\ & \prod_{i=1}^n \left\{ \prod_{j=1}^{\kappa_i^*} \mathbb{I} \left[\phi \left(\omega_{i, \psi_{i,j}^*}^* + (1 - \psi_{i,j}^*/\Delta t_i)x_{i-1} + (\psi_{i,j}^*/\Delta t_i)x_i; \theta \right) < v_{i,j}^* \right] \right\} . \end{aligned}$$

It is now straightforward to integrate out $\Upsilon_i^* = \{v_{i,j}; 1 \leq j \leq \kappa_i^*\}$ and obtain (4.30).

vector of parameters $\theta \in \Theta \subset \mathbf{R}^d$. The process is assumed to be observed without error at a collection of time instances,

$$\{V_{t_0}, V_{t_1}, \dots, V_{t_n}\}, \quad 0 = t_0 < t_1 < \dots < t_n,$$

and the statistical challenge is to explore characteristics, such as the maximum and the level sets, of the likelihood function:

$$\mathfrak{L}(\theta) = \prod_{i=1}^n L_i(\theta), \quad L_i(\theta) := p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta), \quad \theta \in \Theta,$$

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

$$p_t(v, w; \theta) := \mathbf{P}[V_t \in dw \mid V_0 = v; \theta] / dw, \quad t > 0, w, v \in \mathcal{V}. \quad (5.2)$$

is the transition density of (5.1). It is well documented (see for example Sørensen, 2004, for a recent review) that inference about diffusion models is complicated by the unavailability of the transition density (5.2), except for some limited cases. Thus, inference is carried out using either non-likelihood approaches, such as estimating equations (Bibby and Sørensen, 1995), efficient method of moments (Gallant and Tauchen, 1996) and indirect inference (Gourieroux et al., 1993), or approximate likelihood-based approaches, such as computationally intensive MCMC imputation methods (Roberts and Stramer, 2001; Elerian et al., 2001; Eraker, 2001) and methods which approximate analytically the transition density (Aït-Sahalia, 2002). Also, approximate Monte Carlo maximum likelihood approaches have been suggested, most notably by Pedersen (1995a) and Durham and Gallant (2002).

The method described in this paper, termed the Simultaneous Acceptance Method (SAM), estimates each likelihood contribution $L_i(\cdot)$ independently. To simplify the notation, let $L(\cdot)$ denote the likelihood contribution of an arbitrary

Article 5

Monte Carlo Maximum Likelihood Estimation for Discretely Observed Diffusion Processes

5.1 Introduction

This paper introduces a Monte Carlo method for maximum likelihood inference in the context of discretely observed diffusion processes. The method is computationally efficient and gives unbiased and almost surely continuous estimates of the likelihood function, which converge uniformly in the parameters to the likelihood function as the Monte Carlo sample size increases.

Consider scalar time-homogeneous diffusion processes, defined by SDEs of the type:

$$dV_s = b(V_s; \theta) ds + \sigma(V_s; \theta) dW_s, \quad V_s \in \mathcal{V} \subseteq \mathbf{R}, \quad (5.1)$$

where W is a Brownian motion. The functionals $b(\cdot; \theta)$ and $\sigma(\cdot; \theta)$ are called *drift* and *diffusion coefficient* respectively and are assumed to be known up to a

pair of consecutive data points $V_s = v, V_{s+t} = w$, i.e. $L(\theta) = p_t(v, w; \theta)$. SAM generates a *random function* $L(\Xi, \theta)$, $\theta \in \Theta$, where Ξ is a random element *independent* of θ , such that for any fixed $\theta \in \Theta$, $\mathbb{E}[L(\Xi, \theta)] = L(\theta)$, and w.p.1, $\theta \mapsto L(\Xi, \theta)$ is continuous. We estimate $L(\cdot)$ by the Monte Carlo functional averages

$$L^{(K)}(\cdot) := \frac{1}{K} \sum_{k=1}^K L(\Xi_k, \cdot), \quad \Xi_1, \dots, \Xi_K \text{ independent copies of } \Xi. \quad (5.3)$$

Having obtained estimates $L_i^{(K)}(\cdot)$ of $L_i(\cdot)$ using (5.3) independently for each index $i = 1, \dots, n$, we estimate the likelihood function $\mathfrak{L}(\cdot)$ by the product:

$$\mathfrak{L}^{(K)}(\cdot) := \prod_{i=1}^n L_i^{(K)}(\cdot). \quad (5.4)$$

We emphasize that our method yields an estimator of the whole function: conditionally on a fixed-dimensional random element, (5.4) is a *deterministic* function of θ , and critically, the map $\theta \mapsto \mathfrak{L}^{(K)}(\theta)$ is continuous w.p.1.

We demonstrate that our Monte Carlo estimator $\mathfrak{L}^{(K)}(\cdot)$ is computationally very efficient and we detail its theoretical properties. From the computational perspective, the random element consists of standard exponential and gaussian variables and can be easily simulated, and $L(\Xi, \theta)$ is of a simply calculable form. The fact that Ξ is simulated independently of θ allows for the estimation of the likelihood function *simultaneously* for all $\theta \in \Theta$ using a single stream of random numbers. Moreover, the a.s. continuity of $\mathfrak{L}^{(K)}(\cdot)$ facilitates the efficient implementation of optimisation routines for locating its maximiser. From the theoretical perspective, we prove that for any $\theta \in \Theta$, $\mathfrak{L}^{(K)}(\theta)$ is an unbiased estimator of $\mathfrak{L}(\theta)$ with finite moments. More importantly, due to the a.s. continuity of $\mathfrak{L}^{(K)}(\cdot)$, we can resort to the Strong Law of Large Numbers (SLLN) in Banach spaces, to establish that w.p.1 $\mathfrak{L}^{(K)}(\theta)$ converges to $\mathfrak{L}(\theta)$ uniformly in θ , as $K \rightarrow \infty$. Uniform convergence implies, under mild conditions, convergence of the maximisers to the MLE, say

θ_{MLE} :

$$\text{w.p.1, } \lim_{K \rightarrow \infty} \arg \max_{\theta \in \Theta} \mathfrak{L}^{(K)}(\theta) = \theta_{MLE}. \quad (5.5)$$

Also, additional statistics, e.g. profile likelihoods and level sets, can be derived as appropriate limits of the corresponding characteristics of the Monte Carlo estimate of the likelihood.

The construction of the random function $L(\Xi, \cdot)$ is based on a recently developed *retrospective rejection sampling* algorithm called the Exact Algorithm (EA). EA is introduced in Beskos and Roberts (2005) [Article 2] and Beskos et al. (2005c) [Article 3]. It returns a draw from any finite dimensional distribution of the target SDE by rejection sampling with proposals from the Wiener measure. Beskos et al. (2006b) [Article 4] noticed that the transition density of the target diffusion can be written in terms of the acceptance probability of EA and thus devised a simple *pointwise* estimator of $L(\theta)$, termed the Acceptance Method (AM). In this paper AM is evolved into SAM through a series of novel couplings of Brownian motion paths.

The applicability of SAM (and AM) is attached to that of EA, whose latest development (termed EA2 in Beskos et al. (2005c) [Article 3]) covers a class of diffusion processes determined by the set of conditions (C1)–(C3) on the drift and diffusion coefficient given in Section 5.2. Thus, although not universally applicable, SAM can be applied to several diffusion models for which the likelihood function is intractable, e.g. to the class of diffusion processes generated by one-to-one transformations of the linear diffusion process with multiplicative noise (see for example pg.119 of Kloeden and Platen, 1992):

$$dV_s = (\theta_1 + \theta_2 V_s) ds + (\theta_3 + \theta_4 V_s) dW_s, \quad (5.6)$$

with state space $\mathcal{V} = (-\theta_3/\theta_4, \infty)$ or $\mathcal{V} = (-\infty, -\theta_3/\theta_4)$. Undergoing developments in EA (Beskos et al., 2005a) [Article 6] might allow for the extension of the

applicability of SAM to a wider class of models.

The paper is organised as follows. In Section 5.2 we set up the basic notation, provide the random function $L(\Xi, \cdot)$ and state its properties in Theorem 5.1. In Section 5.3 we establish the a.s. uniform convergence of the estimates $\mathfrak{L}^{(K)}(\cdot)$ to $\mathfrak{L}(\cdot)$ and discuss some important consequences. In Section 5.4 we illustrate our method by applying SAM to an example SDE. In Section 5.5 we prove Theorem 5.1 and in Section 5.6 we summarise and discuss possible extensions of the method. The paper is supplemented with a brief Appendix.

5.2 Notation and statement of the main result

In this section we set up the basic notation, construct explicitly the estimator $L(\Xi, \cdot)$ of $L(\cdot) = p_t(v, w; \cdot)$, and state the main result of the paper, which is proven later in Section 5.5. Throughout the rest of the paper we assume that Θ is a compact subset of $\mathbf{R}^d$ which contains θ_{MLE} . Generally, random variables will be written in capital letters; **typewriter** style will be used to emphasise that the distribution of the variable is independent of θ .

We define

$$\eta(u; \theta) = \int^u \frac{1}{\sigma(z; \theta)} dz, \quad u \in \mathcal{V},$$

to be any fixed anti-derivative of $\sigma^{-1}(\cdot; \theta)$. For $u \in \mathbf{R}$, let

$$\alpha(u; \theta) = \frac{b\{\eta^{-1}(u; \theta); \theta\}}{\sigma\{\eta^{-1}(u; \theta); \theta\}} - \sigma'\{\eta^{-1}(u; \theta); \theta\} / 2, \quad A(u; \theta) = \int_0^u \alpha(z; \theta) dz,$$

where η^{-1} is the inverse of η . Throughout the paper we assume that the following conditions hold for all $\theta \in \Theta$:

(C1) $\alpha(\cdot; \theta)$ is continuously differentiable;

(C2) $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded below;

(C3) $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above on (z, ∞) for all $z \in \mathbf{R}$.

We note that if $(\alpha^2 + \alpha')(\cdot; \theta)$ is bounded above on $(-\infty, z)$, $z \in \mathbf{R}$, we redefine $\eta(u; \theta)$, $\alpha(u; \theta)$, as $-\eta(u; \theta)$, and $-\alpha(-u; \theta)$ respectively; then $(\alpha^2 + \alpha')(\cdot; \theta)$ will be bounded above on (z, ∞) , $z \in \mathbf{R}$. We define

$$l(\theta) = \inf_{z \in \mathbf{R}} \frac{(\alpha^2 + \alpha')(z; \theta)}{2}, \quad \phi(u; \theta) = \frac{(\alpha^2 + \alpha')(u; \theta)}{2} - l(\theta) \geq 0,$$

and $r(u; \theta) = \sup_{z \in (u, \infty)} \phi(z; \theta) < \infty$.

The random element is $\Xi = (\mathbf{E}, \Psi, \mathbf{Z}, \mathbf{N})$, where $\mathbf{E} \sim \text{Ex}(1)$, $\mathbf{Z} \sim \mathcal{N}(0, 1)$, Ψ is a homogeneous Poisson process on $[0, t]$ with time-ordered points $\mathbf{Y}_j$, $1 \leq j \leq \Lambda$, $\Lambda \sim \text{Po}(\lambda t)$, where the intensity λ is specified below, and conditionally on Λ , $\mathbf{N} = \{\mathbf{N}_{ij}; 1 \leq i \leq 3, 1 \leq j \leq \Lambda\}$ is a matrix of independent standard gaussian variables. It is understood that on the event $\{\Lambda = 0\}$, Ψ and $\mathbf{N}$ contain no elements. The intensity λ depends on $\mathbf{E}$ and the observed data, v, w , in the following way: we define the functions

$$x = x(\theta) := \eta(v; \theta), \quad y = y(\theta) := \eta(w; \theta), \quad (5.7)$$

$$m = m(\mathbf{E}, \theta) = \left(y + x - \sqrt{2t\mathbf{E} + (y - x)^2} \right) / 2, \quad (5.8)$$

and take λ to be any real not less than $\sup_{\theta \in \Theta} r(m; \theta)$. The following theorem gives the random function $L(\Xi, \cdot)$ as a composition of easily computable functions.

Theorem 5.1 *Assume (C1)–(C3) and the following technical conditions:*

(Itô1) $\eta(\cdot; \theta)$ is twice continuously differentiable for all $\theta \in \Theta$,

(Itô2) $\eta(\mathcal{V}; \theta) = \mathbf{R}$,

(Cnt1) i) $\alpha(\cdot; \cdot), \alpha'(\cdot; \cdot)$, and ii) $A(\cdot; \cdot)$ are continuous on $\mathbf{R} \times \Theta$,

(Cnt2) i) $\eta(u; \cdot)$, ii) $\eta'(u; \cdot)$ are continuous for all $u \in \mathcal{V}$, iii) $l(\cdot)$ is continuous.

Then the random function $L(\Xi, \cdot)$ defined below is such that

i) for any $\theta \in \Theta$, $\mathbb{E}[L(\Xi, \theta)] = L(\theta)$,

ii) w.p.1 $\theta \mapsto L(\Xi, \theta)$ is continuous,

iii) there exists a constant M such that w.p.1 $|L(\Xi, \theta)| < M$ for all $\theta \in \Theta$.

$$L(\Xi, \theta) = |\eta'(w; \theta)| \mathcal{N}_{x,t}(y) \exp\{A(y; \theta) - A(x; \theta) - l(\theta)t\} \times a(\Xi, \theta) ;$$

$$a(\Xi, \theta) = \sum_{i=1}^2 p_i \times \prod_{j=1}^A [1 - \phi(\chi_{ij}; \theta) / \lambda] ;$$

$$\chi_{ij} = m + \sqrt{\left(\beta_{ij} + \sum_{l=1}^j \mathbf{N}_{il} \gamma_{ijl}\right)^2 + \left(\sum_{l=1}^j \mathbf{N}_{2l} \gamma_{ijl}\right)^2 + \left(\sum_{l=1}^j \mathbf{N}_{3l} \gamma_{ijl}\right)^2} ;$$

$$\beta_{ij} = (x - m) \frac{\tau_i - \mathbf{Y}_j}{\tau_i} \mathbb{I}[\mathbf{Y}_j \leq \tau_i] + (y - m) \frac{\mathbf{Y}_j - \tau_i}{t - \tau_i} \mathbb{I}[\mathbf{Y}_j > \tau_i] ;$$

$$\gamma_{ijl}^2 = \frac{(\tau_i - \mathbf{Y}_j)^2 (\mathbf{Y}_l - \mathbf{Y}_{l-1})}{(\tau_i - \mathbf{Y}_l)(\tau_i - \mathbf{Y}_{l-1})} \mathbb{I}[\mathbf{Y}_j \leq \tau_i] + \frac{(t - \mathbf{Y}_j)^2 (\mathbf{Y}_l - \mathbf{Y}_{l-1} \vee \tau_i)}{(t - \mathbf{Y}_l)(t - \mathbf{Y}_{l-1} \vee \tau_i)} \mathbb{I}[\mathbf{Y}_l > \tau_i], \quad \gamma_{ijl} > 0 ;$$

$$\tau_1 = t \left(1 + \frac{y - m}{x - m} g\right)^{-1}, \quad \tau_2 = t \left(1 + \frac{y - m}{x - m} g^{-1}\right)^{-1} ;$$

$$p_1 = \frac{tg\mathbf{E} + 2(x - m)^2}{(1 + g)[t\mathbf{E} + 2(x - m)^2]}, \quad p_2 = 1 - p_1 ;$$

$$g = 1 + \mathbf{Z}^2 / \mathbf{E} - \sqrt{2\mathbf{Z}^2 / \mathbf{E} + \mathbf{Z}^4 / \mathbf{E}^2} > 0 .$$

Remark: In the theorem above, the product $\prod_{j=1}^0$ is defined to be equal to 1.

Moreover, notice that $\chi_{ij}, \beta_{ij}, \gamma_{ijl}, \tau_i, p_1$ are all functions of θ , the random element Ξ and the data, while g depends only on $(\mathbf{E}, \mathbf{Z})$.

Corollary 5.1 *Under the assumptions of Theorem 5.1, the likelihood term $L(\cdot)$ is continuous.*

Proof:

Consider some $\theta \in \Theta$ and a sequence $\{\theta_j\}$ such that $\theta_j \rightarrow \theta$. Then, $\lim_{j \rightarrow \infty} L(\theta_j) =$

$\lim_{j \rightarrow \infty} \mathbb{E}[L(\Xi, \theta_j)] = \mathbb{E}[\lim_{j \rightarrow \infty} L(\Xi, \theta_j)] = L(\theta)$, where the first equality holds due to unbiasedness stated in result i) of Theorem 5.1, the second due to result iii) and bounded convergence and the last due to result ii).

□

5.3 Uniform convergence of the Monte Carlo estimator

$L_i^{(K)}(\theta)$ in (5.3) is an unbiased estimator of $L_i(\theta)$ for any $\theta \in \Theta$, thus Kolmogorov's SLLN implies that $\mathfrak{L}^{(K)}(\theta) \rightarrow \mathfrak{L}(\theta)$ a.s. as $K \rightarrow \infty$. It is known, however, that this pointwise convergence is too weak to guarantee the convergence of the maximisers of $\mathfrak{L}^{(K)}(\cdot)$ to θ_{MLE} , as $K \rightarrow \infty$. Similarly, a stronger form of convergence is needed to ensure that several other interesting features of the Monte Carlo functional averages, such as level sets, integrals over subsets and profile likelihoods, will converge to the corresponding features of the likelihood function. A sufficient condition is that w.p.1 the convergence is uniform in θ :

$$\text{w.p.1, } \lim_{K \rightarrow \infty} \sup_{\theta \in \Theta} |\mathfrak{L}^{(K)}(\theta) - \mathfrak{L}(\theta)| = 0 . \quad (5.9)$$

Questions which lead to essentially equivalent problems to uniform convergence of functional averages occur in theoretical statistics (see for example Burke (1965); Tucker (1959); Talagrand (1987) for the Glivenko-Cantelli problem and Wald (1949); Choirat et al. (2003) for the consistency of the MLE), dynamical systems and ergodic theory, stochastic optimization, econometrics (Andrews, 1987) and Monte Carlo methods (Geyer, 1994); see ch.1 of Peskir (2000) for a review. The general probabilistic framework in which convergence of functional averages is most naturally addressed is *probability on Banach spaces*.

We prove (5.9) using the following fundamental theorem about SLLN for ran-

dom elements in an arbitrary separable Banach space. This result first appeared in Mourier (1953); see also Beck (1963) and Giesy (1976). We recall (see for instance Giesy, 1976) that for an arbitrary random element $\mathcal{X}$ in a Banach space $(\mathfrak{X}, \|\cdot\|)$, $\|\mathcal{X}\|$ denotes its norm, and when $\mathbb{E}[\|\mathcal{X}\|] < \infty$, the expectation $\mathbb{E}[\mathcal{X}]$ is defined as the unique element $\mu \in \mathfrak{X}$ such that $T(\mu) = \mathbb{E}[T(\mathcal{X})]$ for every linear function $T : \mathfrak{X} \mapsto \mathbf{R}$ in the dual space of $\mathfrak{X}$.

Theorem 5.2 *Let $(\mathfrak{X}, \|\cdot\|)$ be a separable Banach space and $\mathcal{X}$ a random element in $\mathfrak{X}$, such that:*

$$\mathbb{E}[\|\mathcal{X}\|] < \infty, \quad \mathbb{E}[\mathcal{X}] = \mathbf{0} .$$

If $\mathcal{X}_1, \mathcal{X}_2, \dots$ are independent copies of $\mathcal{X}$, then:

$$w.p.1, \quad \lim_{K \rightarrow \infty} \left\| \frac{1}{K} \sum_{k=1}^K \mathcal{X}_k \right\| = 0 .$$

The uniform convergence in (5.9) follows easily from the following Corollary.

Corollary 5.2 *Let $L(\Xi, \theta)$ be the random function constructed in Theorem 5.1, and $\Xi_1, \Xi_2, \dots$, independent copies of Ξ . Then:*

$$w.p.1, \quad \lim_{K \rightarrow \infty} \sup_{\theta \in \Theta} \left| \frac{1}{K} \sum_{k=1}^K L(\Xi_k, \theta) - L(\theta) \right| = 0 .$$

Proof:

Take $(\mathfrak{X}, \|\cdot\|)$ to be the space of continuous real functions on the compact set Θ equipped with the sup-norm such that for any $f \in \mathfrak{X}$, $\|f\| = \sup_{\theta \in \Theta} |f(\theta)|$. This is well-known to be a separable Banach space. Theorem 5.1 has shown that $L(\Xi, \cdot)$ takes values in $\mathfrak{X}$, and that $\mathbb{E}[\|L(\Xi, \cdot)\|] < \infty$. Corollary 5.1 has shown that $L(\cdot) \in \mathfrak{X}$, and by uniqueness of expectation $\mathbb{E}[L(\Xi, \cdot) - L(\cdot)] = \mathbf{0}$. Applying Theorem 5.2 to $L(\Xi, \cdot) - L(\cdot)$ yields the result.

□

The compactness of Θ and (5.9) imply the following result, which validates our Monte Carlo maximum likelihood approach.

Corollary 5.3 *Assume that θ_{MLE} is the unique element of $\arg \max_{\theta \in \Theta} \mathfrak{L}(\theta)$. If $\theta_{max}^{(K)} := \arg \max_{\theta \in \Theta} \mathfrak{L}^{(K)}(\theta)$, then w.p.1 $\lim_{K \rightarrow \infty} \theta_{max}^{(K)} = \theta_{MLE}$.*

Note that the same result holds for a sequence of so-called $\epsilon^{(K)}$ -maximisers, i.e. a sequence $\theta_*^{(K)}$ such that $|\theta_{max}^{(K)} - \theta_*^{(K)}| \leq \epsilon^{(K)}$, where $\epsilon^{(K)} \rightarrow 0$ as $K \rightarrow \infty$. This extension allows for numerically efficient implementations of SAM, as in Section 5.4.

Note that a.s. continuity of the random function is not a necessary condition for uniform convergence. For instance, concavity of the functional averages and their expectation would suffice; see Geyer and Thompson (1992). Uniform convergence might hold even when $\Xi_1, \Xi_2, \dots$, form a stationary (not necessarily ergodic) stochastic process; see for example Peskir (2000); Dudley (1999); Choirat et al. (2003); Geyer (1994). On the other hand, consistency results as in Corollary 5.3 could be obtained by establishing epigraphical/hypographical convergence (see for example Attouch, 1984, for relevant results from convex analysis) which is weaker than uniform, although closely related, and requires only semi-continuity properties of $L(\Xi, \cdot)$; see Geyer (1994) and Choirat et al. (2003) for investigations in the context of functional averages. In our case, however, concavity does not necessarily hold, but continuity is rather easily established.

Having shown uniform convergence, we can resort to Theorems 5 and 6 of Geyer (1994) to prove that the Monte Carlo profile log-likelihood and the Monte Carlo level sets converge to the corresponding features of the likelihood function as the number of Monte Carlo samples increases.

5.4 Numerical illustration

We applied SAM to the logistic growth SDE (Beskos et al., 2005c; Rimmer et al., 2005):

$$dV_s = R V_s (1 - V_s/C) ds + \beta V_s dW_s, \quad V_s \in \mathcal{V} = (0, \infty), \quad (5.10)$$

which is used to model the evolution of a population in a environment of capacity C ; R is the rate of growth per individual and $\beta > 0$ a noise parameter. It is known (see for example pg.123 of Kloeden and Platen, 1992) that (5.10) is the inverse of the linear SDE with multiplicative noise (5.6) setting $(\theta_1, \theta_2, \theta_3, \theta_4) = (R/C, \beta^2 - R, 0, \beta)$. Here, we take $\theta = (R, C, \beta)$, and:

$$\eta(u; \theta) = -\log(u)/\beta, \quad \alpha(u; \theta) = \beta/2 - R/\beta + R/(\beta C)e^{-\beta u},$$

$$l(\theta) = \beta^2/8 - R/2, \quad r(u; \theta) = [(\alpha^2 + \alpha')(u; \theta)/2 - l(\theta)] \vee [R^2/(2\beta^2)].$$

If $\Theta = [R_l, R_u] \times [C_l, C_u] \times [\beta_l, \beta_u]$, then for any given pair of successive data points v, w of time increment t ,

$$\lambda = \frac{R_u^2}{2\beta_l^2} \times [(e^{q/2}/C_l - 1)^2 \vee 1], \quad q := \log(vw) + \sqrt{2t\beta_u^2 E + \log^2(w/v)}.$$

It is easy to verify that all conditions of Theorem 5.1 hold.

SAM was applied to a dataset of size $n = 1,000$ simulated using EA (see section 5 of Beskos et al. (2005c) [Article 3] for details) under the specifications $V_0 = 700$, $\theta = (0.1, 1000, 0.1)$ and $\Delta t_i = 1$; the first 250 values are shown in Fig.5.1, left. We chose $\Theta = [0.03, 0.18] \times [850, 1200] \times [0.09, 0.12]$, outside which a preliminary investigation showed that the likelihood is of negligible value. The maximizers of $\mathfrak{L}^{(K)}(\cdot)$, $K = 1, 2, 5, 10, 20, 50, 100, 200, 300$, were found numerically with the downhill simplex method (see for example section 10.4 of Press et al., 1992), up to a precision error $\epsilon^{(K)}$ which decreased with increasing K . The initial search point for each maximization was the output of the previous one (for $K = 1$ the initial

point was $(0.05, 1150, 0.115)$). Computer implementation in C on a Pentium IV 2.6 GHz processor, yielded (in 14 minutes) the sequence of $\epsilon^{(K)}$ -maximisers shown in Fig.5.1, right. The simulated number of Poisson points Λ was on average, over all consecutive data points and Monte Carlo iterations, 2.004 and it never exceeded 11.

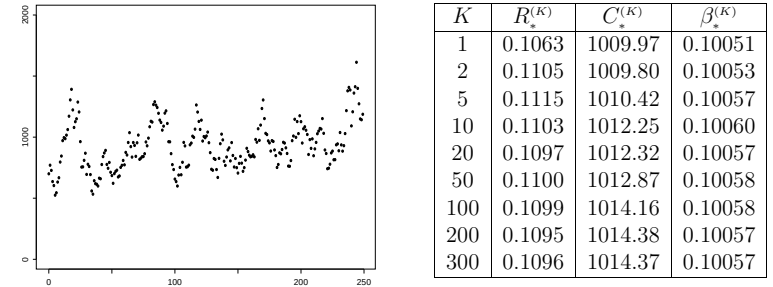


Figure 5.1: Left: The first 250 values of the example dataset from (5.10). Right: Convergence of the $\epsilon^{(K)}$ -maximisers. The dataset was simulated under the parameter values $(0.1, 1000, 0.1)$.

5.5 Proof of Theorem 5.1

The notation in this section follows the definitions in Section 5.2. We prove the result in several stages.

1. Diffusion transformation and densities: We define the unit diffusion coefficient process $X_s := \eta(V_s; \theta)$. Indeed, (Itô1) and (Itô2) allow the application of Itô's lemma to show that:

$$dX_s = \alpha(X_s; \theta) ds + dW_s.$$

Let $\tilde{p}_t(\cdot, \cdot; \theta)$ be the transition density of X defined analogously to (5.2). Recall that we have defined $x = x(\theta) = \eta(v; \theta)$, $y = y(\theta) = \eta(w; \theta)$. A standard change-

of-variables argument yields:

$$L(\theta) \equiv p_t(v, w; \theta) = |\eta'(w; \theta)| \cdot \tilde{p}_t(x, y; \theta), \quad \theta \in \Theta. \quad (5.11)$$

Let $\mathbb{Q}_\theta$ be the law of the paths of X on $[0, t]$ conditioned to begin at $X_0 = x$ and finish at $X_t = y$. We call such a path “a bridge from $(0, x)$ to (t, y) ”. Let $\mathbb{W}_\theta$ be the distribution of the Brownian bridges (BBs) from $(0, x)$ to (t, y) ; a random process distributed according to $\mathbb{W}_\theta$ will be denoted by B . Both $\mathbb{Q}_\theta$ and $\mathbb{W}_\theta$ apply on the space C of continuous mappings from $[0, t]$ to $\mathbf{R}$ equipped with the corresponding cylinder σ -algebra; we denote by ω a typical element of C . Under (Itô2) and (C1) we can use Lemma 1 of Beskos et al. (2006b) [Article 4], which applies Girsanov’s theorem, to find that:

$$\frac{d\mathbb{Q}_\theta}{d\mathbb{W}_\theta}(\omega) = \frac{\mathcal{N}_{x,t}(y)}{\tilde{p}_t(x, y; \theta)} \exp \left\{ A(y; \theta) - A(x; \theta) - \int_0^t \frac{1}{2} (\alpha^2 + \alpha')(\omega_s; \theta) ds \right\}.$$

Taking expectations at both sides w.r.t. $\mathbb{W}_\theta$, using (C2) and re-arranging we get that:

$$\tilde{p}_t(x, y; \theta) = \mathcal{N}_{x,t}(y) \exp \{ A(y; \theta) - A(x; \theta) - l(\theta)t \} \times a(\theta), \quad (5.12)$$

where

$$a(\theta) = \mathbb{E}_{\mathbb{W}_\theta} \left[\exp \left\{ - \int_0^t \phi(\omega_s; \theta) ds \right\} \right] \leq 1.$$

A heuristic description of the remainder of the proof is as follows. We initially derive an unbiased estimator $a(\Xi, \theta)$ of $a(\theta)$, where the distribution of the random element Ξ depends on θ . We then construct jointly the family $\{\Xi; \theta \in \Theta\}$ by expressing $\Xi = f(\Xi, \mathbf{X}, \theta)$, where $\mathbf{X}$ is independent of θ and f is an appropriately specified function. Finally, we show that $a(\Xi, \theta) = \mathbb{E} [a(f(\Xi, \mathbf{X}, \theta), \theta) | \Xi]$, where $\mathbf{X}$ is integrated out to ensure a.s. continuity of the mapping $\theta \mapsto a(\Xi, \theta)$.

2. Connection with exact diffusion bridge simulation: Beskos et al. (2006b) [Article 4] noticed that $a(\theta)$ coincides with the acceptance probability

of a rejection sampling algorithm for the simulation of paths with law $\mathbb{Q}_\theta$. The algorithm, developed in Beskos et al. (2005c) [Article 3] and termed the Exact Algorithm (EA), proposes paths from $\mathbb{W}_\theta$ and accepts them according to the density ratio:

$$\frac{d\mathbb{Q}_\theta}{d\mathbb{W}_\theta}(\omega) \propto \exp \left\{ - \int_0^t \phi(\omega_s; \theta) ds \right\} \leq 1. \quad (5.13)$$

Let m be the minimum of ω , i.e. $m = \inf\{\omega_s; s \in [0, t]\}$; (C3) implies that $\phi(\omega_s; \theta)/r(m; \theta) \leq 1$, for all $s \in [0, t]$. When ω is a realisation of B the marginal distribution of its minimum is known (see for instance Shepp, 1979) and it can be simulated in terms of a standard exponential random variable E according to (5.8); see Proposition 2 of Beskos et al. (2005c) [Article 3]. Notice that if Φ is a Poisson process of intensity $r(m; \theta)$ on $[0, t] \times [0, 1]$ and N the number of the points of Φ lying below the curve $s \mapsto \phi(\omega_s; \theta)/r(m; \theta)$, then

$$P[N = 0 | \omega] = \exp \left\{ - \int_0^t \phi(\omega_s; \theta) ds \right\}. \quad (5.14)$$

Thus, Beskos et al. (2005c) [Article 3] suggested the following simulation algorithm.

Let Ψ be the projection of Φ on the time-axis $[0, t]$ with time-ordered points Y_j , $1 \leq j \leq \Lambda$, where $\Lambda \sim \text{Po}(r(m; \theta)t)$.

Exact Algorithm (EA2)

1. Simulate $E \sim \text{Ex}(1)$, set $m = (y + x - \sqrt{2tE + (y - x)^2}) / 2$.
 2. Simulate Ψ .
 3. Simulate $\{B_{Y_j}; 1 \leq j \leq \Lambda\}$, so that its minimum is m .
 4. With probability $1 - \prod_{j=1}^{\Lambda} [1 - \phi(B_{Y_j}; \theta)/r(m; \theta)]$ goto (1).
 5. Output all information about B .
-

Step 3 requires the simulation of the skeleton $\{B_{Y_j}; 1 \leq j \leq \Lambda\}$ with a given minimum. This can be achieved (see Paragraph 4 below) using existing theory about the decomposition of the Brownian path at its minimum (see for example

Proposition 2 of Asmussen et al., 1995). Therefore, a *pointwise* unbiased estimator of $a(\theta)$ is readily available:

$$a(\Xi, \theta) = \prod_{j=1}^{\Lambda} [1 - \phi(B_{Y_j}; \theta)/r(m; \theta)] . \quad (5.15)$$

where $\Xi = (\mathbf{E}, \Psi, \{B_{Y_j}; 1 \leq j \leq \Lambda\})$. This is the Acceptance Method proposed in Beskos et al. (2006b) [Article 4].

3. Coupling of the Poisson processes: We define the joint structure of the collection of random variables $\{\Psi; \theta \in \Theta\}$ using the thinning property of the Poisson process (see for example section 5 of Kingman, 1993). Let Ψ be a Poisson process on the time axis $[0, t]$ with time-ordered points Y_j , $1 \leq j \leq \Lambda$, of rate $\lambda \geq \sup_{\theta \in \Theta} r(m(\mathbf{E}, \theta); \theta)$, and $\mathbf{U} = (U_1, \dots, U_\Lambda)$ a vector of iid variables, $U_1 \sim \text{Un}(0, 1)$. We set $\Psi = \{Y_j \in \Psi : U_j < r(m; \theta)/\lambda; 1 \leq j \leq \Lambda\}$. Then the right-hand side of (5.15) rewrites as:

$$\prod_{j=1}^{\Lambda} \{1 - \mathbb{I}[U_j < r(m; \theta)/\lambda] \cdot \phi(B_{Y_j}; \theta)/r(m; \theta)\} ,$$

where by integrating out $\mathbf{U}$ we have the following *pointwise* unbiased estimator of $a(\theta)$:

$$a(\Xi, \theta) = \prod_{j=1}^{\Lambda} [1 - \phi(B_{Y_j}; \theta)/\lambda] , \quad (5.16)$$

where now $\Xi = (\mathbf{E}, \Psi, \{B_{Y_j}; 1 \leq j \leq \Lambda\})$.

4. Coupling of the proposed paths: For a path ω with given minimum m , let τ be the instance when m is attained, i.e. $\tau = \sup\{s \in [0, t] : \omega_s = m\}$. As proven in the Appendix, when ω is a realisation of B , the time instance of the minimum can be simulated conditionally on m according to:

$$\tau = \mathbb{I}[\mathbf{V} \leq p_1] \tau_1 + \mathbb{I}[\mathbf{V} > p_1] \tau_2 , \quad (5.17)$$

for $\mathbf{V} \sim \text{Un}(0, 1)$, where we recall that p_1, τ_1, τ_2 are given in terms of $\mathbf{E}, \mathbf{Z}$ and θ . The process B with a given minimum m attained at time τ can be constructed at the time instances Y_j , $1 \leq j \leq \Lambda$, in the following way (see for example section 3.1 of Beskos et al. (2005c) [Article 3]):

$$B_{Y_j} = m + \begin{cases} \sqrt{\tau} \sqrt{\left(\frac{(x-m)(\tau-Y_j)}{\tau^{3/2}} + \mathbf{W}_{1, \frac{Y_j}{\tau}}\right)^2 + \mathbf{W}_{2, \frac{Y_j}{\tau}}^2 + \mathbf{W}_{3, \frac{Y_j}{\tau}}^2}, & \text{if } Y_j \leq \tau , \\ \sqrt{t-\tau} \sqrt{\left(\frac{(y-m)(Y_j-\tau)}{(t-\tau)^{3/2}} + \mathbf{W}_{4, \frac{Y_j-\tau}{t-\tau}}\right)^2 + \mathbf{W}_{5, \frac{Y_j-\tau}{t-\tau}}^2 + \mathbf{W}_{6, \frac{Y_j-\tau}{t-\tau}}^2}, & \text{if } Y_j > \tau , \end{cases} \quad (5.18)$$

where $\mathbf{W} = \{\mathbf{W}_i; 1 \leq i \leq 6\}$, is a collection of six independent BBs from $(0, 0)$ to $(1, 0)$. We define B_{i, Y_j} as B_{Y_j} in (5.18) under the specification $\tau = \tau_i$, for $i = 1, 2$ and $1 \leq j \leq \Lambda$. Under this convention, substituting (5.17), (5.18) into (5.16) and integrating out $\mathbf{V}$, we re-write the function we wish to estimate as:

$$\begin{aligned} a(\theta) &= \mathbb{E} \left\{ \mathbb{E} \left[\prod_{j=1}^{\Lambda} [1 - \phi(B_{Y_j}; \theta)/\lambda] \mid \mathbf{E}, \mathbf{Z}, \mathbf{W}, \Psi \right] \right\} = \\ &= \sum_{i=1}^2 \mathbb{E} \left\{ \prod_{j=1}^{\Lambda} [1 - \phi(B_{i, Y_j}; \theta)/\lambda] \times p_i \right\} = a_1(\theta) + a_2(\theta) , \end{aligned}$$

where $a_1(\theta)$, $a_2(\theta)$ are defined in the obvious way. In the sequel, we construct estimators $a_i(\Xi, \theta)$ of $a_i(\theta)$, $i = 1, 2$, and take $a(\Xi, \theta) = a_1(\Xi, \theta) + a_2(\Xi, \theta)$.

For each Y_j , we need to simulate the bridges $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3$ at the instance Y_j/τ_i or simulate $\mathbf{W}_4, \mathbf{W}_5, \mathbf{W}_6$ at the instance $(Y_j - \tau_i)/(t - \tau_i)$ if $Y_j \leq \tau_i$ or $Y_j > \tau_i$ respectively, $1 \leq j \leq \Lambda$; see also Fig.5.2 for a graphical illustration. Thus, in total we need $3 \times \Lambda$ simulations, which can be done using the same number of gaussian random variables as we now show.

We can simulate a standard BB, $\{\mathbf{W}_s; s \in [0, 1]\}$, at some given time instances $0 < s_1 < \dots < s_d < 1$ using the formula (derived using standard properties of

BB):

$$W_{s_j} = \sum_{l=1}^j N_l \sqrt{\frac{(1-s_j)^2(s_l - s_{l-1})}{(1-s_{l-1})(1-s_l)}}, \quad 1 \leq j \leq d, \quad (5.19)$$

where $N_1, N_2, \dots, N_d$ are iid with $N_1 \sim \mathcal{N}(0, 1)$.

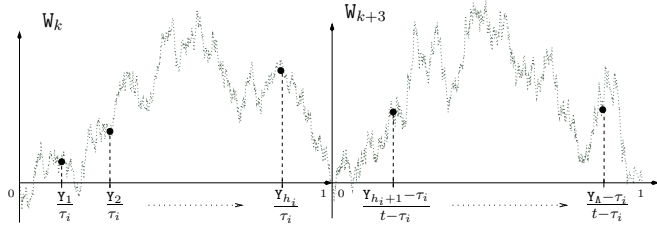


Figure 5.2: Scheme for simulating W_k and W_{k+3} , $k = 1, 2, 3$, at the required discrete time instances; we have defined $h_i = \sum_{j=1}^{\Lambda} \mathbb{I}[Y_j \leq \tau_i]$.

Let $\mathbf{N} = \{N_{kj}; 1 \leq k \leq 3, 1 \leq j \leq \Lambda\}$ be a collection of independent $\mathcal{N}(0, 1)$ variables. We will use $\{N_{kj}; 1 \leq j \leq \Lambda\}$ for the generation of the required instances of the bridges W_k and W_{k+3} , $1 \leq k \leq 3$, see Fig.5.2. Following (5.19) we obtain:

$$W_{k, \frac{Y_j}{\tau_i}} = \sum_{l=1}^j N_{kl} \sqrt{\frac{(\tau_i - Y_j)^2(Y_l - Y_{l-1})}{\tau_i(\tau_i - Y_l)(\tau_i - Y_{l-1})}}, \quad Y_j \leq \tau_i, \\ W_{k+3, \frac{Y_j - \tau_i}{t - \tau_i}} = \sum_{l=1}^j N_{kl} \sqrt{\frac{(\tau_i - Y_j)^2(Y_l - Y_{l-1} \vee \tau_i)}{(t - \tau_i)(\tau_i - Y_l)(\tau_i - Y_{l-1} \vee \tau_i)}}, \quad \mathbb{I}[Y_l > \tau_i], \quad Y_j > \tau_i,$$

for $1 \leq j \leq \Lambda$, $k = 1, 2, 3$, and for both $i = 1, 2$. These expressions allow us to write the proposed paths as deterministic functions of $\mathbf{E}, \Psi, \mathbf{Z}$ and $\mathbf{N}$. In particular, substituting the above expressions into (5.18) and setting χ_{ij} equal to the right-hand side of (5.18) for $\tau = \tau_i$, $i = 1, 2$, $1 \leq j \leq \Lambda$, we take

$$a_i(\Xi, \theta) = p_i \times \prod_{j=1}^{\Lambda} [1 - \phi(\chi_{ij}; \theta) / \lambda], \quad i = 1, 2.$$

5. A.s. continuity of the random function: (Cnt2).i implies that β_{ij}, γ_{ijl} are continuous functions of θ for all i, j, l (even at values of θ such that $Y_j = \tau_i$

or $Y_l = \tau_i$); so, χ_{ij} is continuous in θ . (Cnt1).i, (Cnt2).iii suggest that $\phi(\cdot; \cdot)$ is continuous on $\mathbf{R} \times \Theta$, thus $\phi(\chi_{ij}; \theta)$ and $a(\Xi, \theta)$ are continuous in θ . The continuity of $L(\Xi, \theta)$ follows from Assumptions (Cnt1).ii and (Cnt2).

6. A.s. boundedness of the random function: By definition $a(\Xi, \theta) \leq 1$ a.s., thus w.p.1 $|L(\Xi, \theta)| < M$ for all $\theta \in \Theta$ where:

$$M = \sup_{\theta \in \Theta} \left\{ |\eta'(w; \theta)| \mathcal{N}_{x,t}(y) \exp\{A(y; \theta) - A(x; \theta) - l(\theta)t\} \right\} < \infty.$$

M is finite as the supremum of a continuous function over a compact set.

5.6 Discussion

In this paper we have introduced a new method for likelihood inference for discretely observed diffusions. The method is highly efficient and simple to implement, and expands significantly the family of diffusion models for which routine maximum likelihood calculations are possible. Applications of the approach advocated here together with other likelihood methods based on EA are given in Beskos et al. (2006b) [Article 4].

Our methodology is a Monte Carlo approach based on two types of probabilistic constructions. The first of these exploits a duality between diffusions and Poisson processes, which has become transparent since the development of EA in Beskos et al. (2005c) [Article 3]. The second, involves explicit representations of conditioned Brownian sample paths constructed so as to be suitably continuous in model parameters. Although the entire mathematical construction is quite involved, in this paper we are able to distill its implementation down to a very simple Monte Carlo algorithm requiring merely collections of gaussian and exponential random seeds, described in Theorem 5.1.

The method gives unbiased and continuous estimates of likelihood functions. The unbiasedness of our estimator is in contrast with many competing approaches

which involve biases due to time-discretization errors. The continuity of the estimator, in conjunction with the simplicity of our Monte Carlo algorithms, permit us to demonstrate a rather general LLN which implies (among other properties) that Monte Carlo MLEs converge a.s. to the true MLE. A natural by-product of the approach is that likelihood estimators are bounded so that all moments are bounded and estimates are very stable. The advantages of our approach are supported also by numerical evidence in Section 5.4, where we estimate the parameters of the logistic growth SDE.

The scope of the methodology we introduce here is likely to be extended by currently ongoing work on simulating diffusions for which condition (C3) is not required to hold.

APPENDIX

Simulation of τ :

Proposition 2 of Beskos et al. (2005c) [Article 3] implies that the marginal distribution of the time instance that the minimum m of B is attained can be simulated according to:

$$\tau^{-1} = \left\{ 1 + \mathbb{I} \left[U_0 < (1 + \sqrt{c_1/c_2})^{-1} \right] \cdot I_1 + \mathbb{I} \left[U_0 \geq (1 + \sqrt{c_1/c_2})^{-1} \right] \cdot I_2^{-1} \right\} / t ,$$

where $c_1 = (y - m)^2/(2t)$, $c_2 = (x - m)^2/(2t)$, and $I_1 \sim IGau(\sqrt{c_1/c_2}, 2c_1)$, $I_2 \sim IGau(\sqrt{c_2/c_1}, 2c_2)$ and $U_0 \sim \text{Un}(0, 1)$. We denote by $IGau(\cdot, \cdot)$ the inverse gaussian distribution specified by two positive parameters (see for instance ch.IV of Devroye, 1986). An inverse gaussian random variable I with parameters (c, d) can be represented as (see pg.149 of Devroye, 1986):

$$I = \mathbb{I} \left[V_0 \leq 1/(1+z) \right] cz + \mathbb{I} \left[V_0 > 1/(1+z) \right] \frac{c}{z} ,$$

where $z = 1 + (c/d)Z^2/2 - \sqrt{4(c/d)Z^2 + (c/d)^2Z^4}/2$, $V_0 \sim \text{Un}(0, 1)$. Using this formula for I_1, I_2 with the same V_0, Z , and after some algebra, the expression for τ simplifies to (5.17).

Article 6

A Factorisation of Diffusion Measure and Finite Sample Path Constructions

6.1 Introduction

In this paper, we shall give explicit and a.s. finite constructions for diffusion processes. Our methods give rise to efficient algorithms which can simulate exactly (i.e. free of any time discretisation error and subject only to finite computer precision) from any finite-dimensional distribution of the diffusion. Moreover, we devise an algorithm for the simulation of diffusion hitting times of smooth boundaries. The methodology presented in this paper is based on a novel construction of the Brownian bridge with known range for its extrema.

For ease of exposition, we commence by considering the one-dimensional diffusion process $X = \{X_s; s \in [0, t]\}$ defined through the SDE:

$$dX_s = \alpha(X_s) ds + dW_s, \quad s \in [0, t], \quad X_0 = x, \quad (6.1)$$

where $W = \{W_s; s \in [0, t]\}$ is a scalar Brownian motion and $\alpha : \mathbf{R} \mapsto \mathbf{R}$ some drift

function. In later sections we will discuss generalizations to time-inhomogeneous and multivariate processes.

Assume that (6.1) admits a unique weak and non-explosive solution. Let $\mathbb{Q}$ and $\mathbb{W}$ denote the law of X and the Wiener measure respectively on $[0, t]$ for the initial condition $X_0 = W_0 = x$. In the rest of the paper we assume the following conditions.

(C0) The Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{W}$ exists and it is given by Girsanov's formula,

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp \left\{ \int_0^t \alpha(X_s) dX_s - \int_0^t \frac{1}{2} \alpha^2(X_s) ds \right\}.$$

(C1) $\alpha \in C^1(\mathbf{R})$,

(C2) $\alpha^2 + \alpha'$ is bounded below.

The aim is to perform rejection sampling from $\mathbb{Q}$ using candidates from $\mathbb{W}$. Under (C1), we can use Itô's lemma to rewrite Girsanov's formula as

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp \left\{ A(X_t) - A(x) - \int_0^t \frac{\alpha^2(X_s) + \alpha'(X_s)}{2} ds \right\}, \quad (6.2)$$

for

$$A(u) = \int_0^u \alpha(z) dz, \quad u \in \mathbf{R}.$$

The theory to be presented extends the work of Beskos et al. (2005c) [Article 3] which provides simulation algorithms under the additional condition:

$$(*) \limsup_{u \rightarrow \infty} (\alpha^2 + \alpha')(u) < \infty, \text{ or } \limsup_{u \rightarrow -\infty} (\alpha^2 + \alpha')(u) < \infty.$$

Whilst (C1), (C2) will be naturally satisfied in many contexts, (*) is certainly restrictive (for example is not met in the Ornstein-Uhlenbeck case). In this paper we remove (*). With small modifications, the resulted algorithm can also provide ex-

act draws from the distribution of the hitting time of one-sided smooth boundaries for the Brownian motion and other diffusions.

The paper is organised as follows. In Section 6.2 we introduce the diffusion law factorization which gives rise to the explicit diffusion construction, and we provide two classes of processes for which this factorization has already been used for their exact simulation. Section 6.3 then gives the general construction for one-dimensional homogeneous diffusions. This is generalised in Section 6.4 to multivariate and time-inhomogeneous diffusions. Section 6.5 gives an explicit construction for the hitting time of Brownian motion to an arbitrary smooth boundary. This construction extends naturally to arbitrary diffusions. Section 6.6 concludes with some remarks about possible extensions and future work.

6.2 The Wiener-Poisson factorisation of $\mathbb{Q}$

Additionally to (C0)–(C2), we assume that the following condition holds for the rest of the paper:

(C3) The function $h(u) = \exp\{-(u-x)^2/(2t) + A(u)\}$ is integrable.

This is a weak assumption, since for example it is satisfied (at least for sufficiently small t) when a linear growth bound is assumed for the drift. Then, we define the probability measure $\mathbb{Z}$ through its Radon-Nikodym derivative with respect to $\mathbb{W}$:

$$\frac{d\mathbb{Z}}{d\mathbb{W}} \propto \exp\{A(X_t) - A(x)\} . \quad (6.3)$$

We can interpret the law $\mathbb{Z}$ as a mixture of Brownian bridges weighted by the quantity $h(X_t)/\int_{\mathbf{R}} h(u)du$.

Let $\mathbb{L}$ denote the law of a unit rate Poisson process on $[0, t] \times [0, \infty)$, and define the extended law $\mathbb{Z} \otimes \mathbb{L}$ with typical realisation (X, Φ) . When necessary, we write $\Phi = \{(\psi_j, v_j)\}_{j \geq 1}$, with $\{v_j\}$ non-decreasing. We define the non-negative drift

functional,

$$\phi(u) := \frac{\alpha^2(u) + \alpha'(u)}{2} - l, \text{ where } l = \inf_{z \in \mathbf{R}} \frac{\alpha^2(z) + \alpha'(z)}{2} ,$$

and consider the epigraph of $s \mapsto \phi(X_s)$,

$$\text{epi } \phi(X) := \{ (s, u) \in [0, t] \times \mathbf{R}_+ : \phi(X_s) \leq u \} ,$$

and the event

$$\Gamma := \{\Phi \subset \text{epi } \phi(X)\} = \bigcap_{j \geq 1} \{\phi(X_{\psi_j}) < v_j\} .$$

Theorem 6.1 (Wiener-Poisson factorization) $\mathbb{Q}$ is the marginal distribution of X when $(X, \Phi) \sim \mathbb{Z} \otimes \mathbb{L} \mid \Gamma$.

Proof:

(6.2) and (6.3) yield that

$$\frac{d\mathbb{Q}}{d\mathbb{Z}} \propto \exp\left\{-\int_0^t \phi(X_s)ds\right\} .$$

From the explicit calculation of the zero probability of a Poisson random variable:

$$\mathbb{L}(\Gamma \mid X) = \exp\left\{-\int_0^t \phi(X_s)ds\right\} .$$

Thus the result holds by Bayes's theorem. □

Theorem 6.1 suggests a rejection sampling for the simulation of sample paths from $\mathbb{Q}$: propose X from $\mathbb{Z}$, simulate Φ , and if Γ has occurred then return X . This algorithm can be seen as a construction on the space $C[0, t] \times ([0, t] \times [0, \infty))^{\mathbf{N}}$. However, the aim of this paper is to provide exact but *finite* constructions of X . This can be done, under the following general scheme. We construct a finite-

dimensional random variable $\mathcal{X}$, and a positive function r with the properties that:

1. $r(\mathcal{X}) < \infty$ a.s.,
2. $\sup_{s \in [0, t]} \phi(X_s) \leq r(\mathcal{X})$ a.s..

In Section 6.5 we will allow $\mathcal{X}$ to be constructed on an enlargement of the canonical probability space for X . However, for immediate application we shall just use $\mathcal{X} = \mathcal{X}(X)$. Hence,

$$\Gamma = \Gamma_{r(\mathcal{X})} := \bigcap_{j: v_j < r(\mathcal{X})} \{\phi(X_{\psi_j}) < v_j\}.$$

We define

$$\Phi_{r(\mathcal{X})} = \Phi \cap ([0, t] \times [0, r(\mathcal{X})]).$$

Exact Algorithm (EA)

1. **Simulate** $X_t \sim h$.
 2. **Simulate** $\mathcal{X}$ conditionally on X_t .
 3. **Simulate** $\Phi_{r(\mathcal{X})}$.
 4. **Simulate** $\{X_{\psi_i}; 1 \leq i \leq |\Phi_{r(\mathcal{X})}|\}$ from $\mathbb{Z} | X_t, \mathcal{X}$.
 5. **If** $\Gamma_{r(\mathcal{X})}$ **has occurred** **output** $\{X_{\psi_i}; 1 \leq i \leq |\Phi_{r(\mathcal{X})}|\}$, **otherwise** **goto** (1).
-

When $\Gamma_{r(\mathcal{X})}$ has occurred, $\{X_{\psi_i}; 1 \leq i \leq |\Phi_{r(\mathcal{X})}|\} \sim \mathbb{Q}$. Thus, the only remaining challenge is to characterise the conditional distributions at Steps 2 and Step 4 of the algorithm. The two steps are carried out conditionally on the ending point X_t . From the definition of $\mathbb{Z}$, this implies that we will actually work with Brownian bridge dynamics.

6.2.1 Exact Algorithm 1 (EA1)

In Beskos and Roberts (2005) [Article 2] the Exact Algorithm (EA) was developed under the strong condition

(**) the function ϕ is bounded above.

In this simplified setting, additional information about the proposed path is not required to determine the restricted Poisson process, since it suffices to take

$$r(u) = r = \sup_{z \in \mathbf{R}} \phi(z).$$

Thus, Step 4 merely involves Brownian bridge simulation. We define the skeleton

$$\mathcal{S} = \{(0, X_0), (t, X_t)\} \cup \{(\psi_i, X_{\psi_i}); 1 \leq i \leq |\Phi_r|\}.$$

An important feature of the method, stated in Theorem 6.2 below, is that the distribution of $X | \mathcal{S}, \Gamma$ is derived solely from the proposal measure $\mathbb{Z}$. We denote by $\mathbb{B}(s_1, s_2; u_1, u_2)$ the law of a Brownian bridge starting at time s_1 at u_1 and ending at time s_2 at u_2 . When $s_1 = 0$ we write $\mathbb{B}(s_2; u_1, u_2)$; we simplify the notation to $\mathbb{B}$ when the parameters specifying the bridge are clearly understood from the context.

Theorem 6.2 Under (C0)–(C3) and (**),

$$X | \mathcal{S}, \Gamma \sim \bigotimes_{i=1}^{|\Phi_r|+1} \mathbb{B}(s_{i-1}, s_i; X_{s_{i-1}}, X_{s_i})$$

where $\{s_i; 0 \leq i \leq |\Phi_r| + 1\}$ is an increasing ordering of the set of time instances $\{0, t\} \cup \{\psi_i; 1 \leq i \leq |\Phi_r|\}$.

Proof:

Notice that $X | \mathcal{S}, \Gamma \stackrel{d}{=} X | \mathcal{S}$.

□

6.2.2 EA2 and decomposition of the Brownian path at its minimum

Beskos et al. (2005c) replaced condition (**) with the weaker (*). Without loss of generality, assume that $\limsup_{u \rightarrow \infty} (\alpha^2 + \alpha')(u) < \infty$, so the function ϕ is bounded above on $[u, \infty)$ for any real u . Let $m = \inf\{X_s; s \in [0, t]\}$. It suffices to set

$$\mathcal{X} = m, \quad \text{and} \quad r(u) = \sup_{z \in [u, \infty)} \phi(z), \quad u \in \mathbf{R}.$$

The law of the minimum of a Brownian bridge is known (Shepp, 1979) and it can be easily simulated using the inverse CDF method (see Proposition 2 of Beskos et al. (2005c) [Article 3]). Step 4 requires the simulation of a Brownian bridge with a given minimum. This conditioned process can be represented in terms of two independent Bessel bridges (see for example Asmussen et al., 1995). The construction proceeds as follows.

We will need the instance of the minimum $\tau = \sup\{s \in [0, t] : X_s = m\}$. We denote by $\mathbb{S}(s_1, s_2; u_1, u_2)$ the law of a (3-dimensional) Bessel bridge starting at time s_1 at $u_1 \geq 0$ and finishing at time s_2 at $u_2 \geq 0$. We simplify to $\mathbb{S}(s_2; u_1, u_2)$ when $s_1 = 0$. Assume that $X_t = y$. Let $R_1 \sim \mathbb{S}(\tau; x - m, 0)$ and $R_2 \sim \mathbb{S}(t - \tau; 0, y - m)$ independently. The process

$$m + R_{s,1} \mathbb{I}[0 \leq s < \tau] + R_{s-\tau,2} \mathbb{I}[\tau \leq s \leq t], \quad s \in [0, t],$$

has law $\mathbb{BB}(t; x, y)$. Simulation of τ conditionally on m can be carried out following the algorithm given in Beskos et al. (2005c) [Article 3]. Also, Bessel bridges starting or finishing at 0 can be easily constructed in terms of three independent Brownian bridges. The explicit formula is stated in Section 6.5.

In this setting,

$$\mathcal{S} = \{(0, X_0), (t, X_t), (\tau, X_\tau)\} \cup \{(\psi_i, X_{\psi_i}); 1 \leq i \leq |\Phi_{r(\mathcal{X})}|\}.$$

As in EA1, the conditional law of X given all simulated variables is easily identified.

Theorem 6.3 *Under (C0)–(C3), and (*),*

$$X - \mathcal{X} | \mathcal{S}, \mathcal{X}, \Gamma \sim \bigotimes_{i=1}^{|\Phi_{r(\mathcal{X})}|+2} \mathbb{S}(s_{i-1}, s_i; X_{s_{i-1}} - \mathcal{X}, X_{s_i} - \mathcal{X})$$

where $\{s_i; 0 \leq i \leq |\Phi_{r(\mathcal{X})}| + 2\}$ is an increasing ordering of the time instances $\{0, t, \tau\} \cup \{\psi_i; 1 \leq i \leq |\Phi_{r(\mathcal{X})}|\}$.

6.3 Exact Algorithm 3 (EA3)

In this section we demonstrate how to apply EA when only (C0)–(C3) are assumed. Let $\{a_i\}_{i \geq 1}$ and $\{b_i\}_{i \geq 1}$ be two increasing sequence of positive real numbers, and $a_0 = b_0 = 0$. Given $X_0 = x$, $X_t = y$, we set $\bar{x} = x \wedge y$, $\bar{y} = x \vee y$, and define the following events:

$$U_i = \left\{ \sup_{0 \leq s \leq t} X_s \in [\bar{y} + b_{i-1}, \bar{y} + b_i] \right\} \cap \left\{ \inf_{0 \leq s \leq t} X_s > \bar{x} - a_i \right\},$$

$$L_i = \left\{ \inf_{0 \leq s \leq t} X_s \in (\bar{x} - a_i, \bar{x} - a_{i-1}] \right\} \cap \left\{ \sup_{0 \leq s \leq t} X_s < \bar{y} + b_i \right\},$$

and

$$D_i = U_i \cup L_i, \quad i \geq 1.$$

The sets $D_i, i \geq 1$, constitute a partition of the sample space. We introduce the discrete random variable $I = I(X)$ such that $\{I = i\} = D_i$. Note that $\{I \leq i\}$ is

equivalent to $\{\bar{x} - a_i < X_s < \bar{y} + b_i, \text{ for all } s \in [0, t]\}$. Thus, we set

$$\mathcal{X} = I, \quad \text{and} \quad r(i) = \sup \{ \phi(z); z \in (\bar{x} - a_i, \bar{y} + b_i) \}, \quad i \in \mathbf{N}.$$

Fig.6.1 demonstrates the construction.

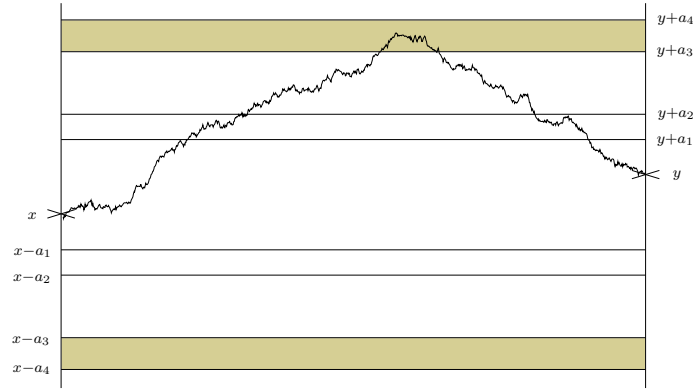


Figure 6.1: A sample path from $X_0 = x$ to $X_t = y$, with $x < y$. In this example we have used symmetric layers, $a_i = b_i, i \geq 1$. For the specific choice of $\{a_i\}$ illustrated in the graph, the event U_4 has occurred. Thus, in this case $I(X) = 4$.

In the remainder of the section we exploit a representation of the CDF of I in terms of monotone alternating series, to develop a novel algorithm for its simulation. We then show how to simulate the Brownian bridge conditioned on I . In the sequel, we will sometimes call I the layer of the Brownian bridge. Accordingly, we refer to a Brownian bridge conditioned on some value of I as *layered* Brownian bridge. The algorithm devised for the simulation of I can be readily modified to develop a rejection sampler of the layered Brownian bridge using Brownian bridge proposals. However, the number of proposed paths until the first acceptance has infinite expectation, where the expectation is taken w.r.t. proposed paths and I . This problem is alleviated by introducing an alternative simulation-driven construction of the layered Brownian bridge, which requires choosing $a_i = b_i, i \geq 1$.

6.3.1 Some auxiliary results

In the following subsections we identify events relevant to the construction of the layered Brownian path whose probability, p say, is not known explicitly but it can be expressed as the limit of the following series of over/under-estimations:

$$0 < S_2 < S_4 < S_6 < \dots < p < \dots < S_5 < S_3 < S_1,$$

for some explicitly known $\{S_j\}$. To simplify exposition we consider the following definition.

Definition 6.1 Let $\mathfrak{T}$ be the subset of Cauchy sequences with values in $\mathbf{R}_+$, such that $\{S_j\} \in \mathfrak{T}$ if and only if

$$S_{2j} < S_{2j+2} < S_{2j+1} < S_{2j-1}, \text{ for all } j \geq 1.$$

For our constructions we will require the simulation of events having probability $p = \lim_{j \rightarrow \infty} S_j$, where $\{S_j\} \in \mathfrak{T}$. Thus, we need to simulate the binary random variable $\mathbb{I}[U < p]$, where $U \sim \text{Un}(0, 1)$. This is accomplished in the following way. Let

$$J = \inf\{j : j \text{ even, } S_j > U \text{ or } j \text{ odd, } S_j < U\}.$$

Then

$$\mathbb{I}[U < p] = \mathbb{I}[J \text{ is even}].$$

Thus, one needs to consider only J terms of $\{S_j\}$ to simulate the event of probability p . The specific alternating series that appear in our algorithms and are presented in the sequel, are converging in their limit faster than exponentially, so J will be typically of a very small value.

Before proceeding, we note the following property which simplifies greatly the simulation procedures we will describe.

Proposition 6.1 *Let $f : \mathbf{R}_+^m \rightarrow \mathbf{R}_+$, for some $m \in \mathbf{N}$, be a C^1 function with*

$$\left| \frac{\partial f}{\partial u_i}(\mathbf{u}) \right| > 0 \text{ for all } 1 \leq i \leq m, \text{ and } \mathbf{u} \in \mathbf{R}_+^m.$$

Let $\{S_j^i\} \in \mathfrak{T}$, for all $1 \leq i \leq m$, with $p_i = \lim_{j \rightarrow \infty} S_j^i$, and set $p = f(p_1, \dots, p_m)$. Consider the sequence $\{S_j\}$, where S_j equals f evaluated at its i -th argument at S_j^i (resp. S_{j+1}^i) when it is increasing (resp. decreasing) w.r.t. that argument. Then, $\{S_j\} \in \mathfrak{T}$ and $p = \lim_{j \rightarrow \infty} S_j$.

The absolute maximum of a Brownian bridge

For $K > |u_1| \vee |u_2|$, let $\gamma(s, u_1, u_2, K)$ denote the probability under $\mathbb{B}\mathbb{B}(s; u_1, u_2)$ that a path does not leave the interval $[-K, K]$. This probability has a known expression in terms of an infinite series, a result which can be traced back to Doob (1949), although see Pötzelberger and Wang (2001) for a more recent reference. We define for $j \geq 1$,

$$\begin{aligned} \bar{\sigma}_j(s, u_1, u_2, K) &= \exp \left\{ -\frac{2}{s} [2Kj - (K + u_1)] [2Kj - (K + u_2)] \right\}, \\ \bar{\tau}_j(s, u_1, u_2, K) &= \exp \left\{ -\frac{2j}{s} [4K^2j + 2K(u_1 - u_2)] \right\}. \end{aligned}$$

Then, Theorem 3 of Pötzelberger and Wang (2001) yields

$$\gamma(s, u_1, u_2, K) = 1 - \sum_{j=1}^{\infty} \{ \sigma_j(s, u_1, u_2, K) - \tau_j(s, u_1, u_2, K) \}, \quad (6.4)$$

where

$$\sigma_j = \bar{\sigma}_j(s, u_1, u_2, K) + \bar{\sigma}_j(s, -u_1, -u_2, K), \quad \tau_j = \bar{\tau}_j(s, u_1, u_2, K) + \bar{\tau}_j(s, -u_1, -u_2, K).$$

Notice that due to symmetry properties of the Brownian bridge, (6.4) can be used to compute the probability that an arbitrary Brownian bridge does not escape

any given interval. Simulation of events of probability (6.4) can be achieved as described earlier due to Proposition 6.2 below.

Proposition 6.2 *Let $\{S_j\}$ be constructed as follows:*

$$S_{2j-1} = \sum_{k=1}^{j-1} (\sigma_k - \tau_k) + \sigma_j, \quad S_{2j} = S_{2j-1} - \tau_j, \quad j \geq 1,$$

where $\sigma_j = \sigma_j(s, u_1, u_2, K)$, $\tau_j = \tau_j(s, u_1, u_2, K)$. Then $\{S_j\} \in \mathfrak{T}$ for any values $u_1, u_2 \in \mathbf{R}$, $s > 0$, and $K > |u_1| \vee |u_2|$.

Proof:

Routine calculation show $\bar{\sigma}_j > \bar{\tau}_j > \bar{\sigma}_{j+1}$, for all $j \geq 1$. The result follows directly. $\square$

Hitting probabilities of the Bessel Bridge

Consider some $u_1 \geq 0$, $u_2 > 0$ and $K > u_1 \vee u_2$. Let $\delta(s, u_1, u_2, K)$ denote the probability under the Bessel bridge law $\mathbb{S}(s; u_1, u_2)$ that a path does not exceed K . Similarly, for $K < L$ we let $\delta(s, u_1, u_2, K; L)$ denote the probability under $\mathbb{S}(s; u_1, u_2)$ that a path conditioned to remain below L , does not exceed K . The derivation of these probabilities exploits the representation of a Bessel bridge as a Brownian bridge conditioned to stay positive. Thus, for $u_1 > 0$ and omitting the arguments of the Brownian bridge measure,

$$\delta(s, u_1, u_2, K; L) = \frac{\mathbb{B}\mathbb{B}[X \text{ does not leave } (0, K)]}{\mathbb{B}\mathbb{B}[X \text{ does not leave } (0, L)]} = \frac{\gamma(s, u_1 - K/2, u_2 - K/2, K/2)}{\gamma(s, u_1 - L/2, u_2 - L/2, L/2)},$$

$$\delta(s, u_1, u_2, K) = \lim_{L \rightarrow \infty} \delta(s, u_1, u_2, K; L) = \frac{\gamma(s, u_1 - K/2, u_2 - K/2, K/2)}{1 - \exp\{-2u_1u_2/s\}}.$$

In this setting, $\mathbb{B} \equiv \mathbb{B}(s; u_1, u_2)$. For the special case $u_1 = 0$, we take the limit $u_1 \rightarrow 0$ in the above expressions and find that,

$$\delta(s, 0, u_2, K; L) = \frac{u_2 - \sum_{j=1}^{\infty} \{\zeta_j(s, u_2, K) - \xi_j(s, u_2, K)\}}{u_2 - \sum_{j=1}^{\infty} \{\zeta_j(s, u_2, L) - \xi_j(s, u_2, L)\}} ,$$

$$\delta(s, 0, u_2, K) = 1 - \frac{1}{u_2} \sum_{j=1}^{\infty} \{\zeta_j(s, u_2, K) - \xi_j(s, u_2, K)\} ,$$

where

$$\xi_j(s, u_2, K) = (2Kj + u_2) \exp \{-2Kj(Kj + u_2)/s\} , \quad \zeta_j(s, u_2, K) = \xi_j(s, -u_2, K) .$$

Simulation of events of probability $\delta(s, u_1, u_2, K; L)$ and $\delta(s, u_1, u_2, K)$ can be easily achieved for any $u_1, u_2 > 0$, $L > K > u_1 \vee u_2$, exploiting Proposition 6.1. For the limiting case, $u_1 = 0$, we additionally require the following result.

Proposition 6.3 *Let $\{S_j\}$ be constructed as follows:*

$$S_{2j-1} = \sum_{k=1}^{j-1} (\zeta_k - \xi_k) + \zeta_j, \quad S_{2j} = S_{2j-1} - \xi_j, \quad j \geq 1,$$

where $\zeta_j = \zeta_j(s, u_2, K)$, $\xi_j = \xi_j(s, u_2, K)$, with $K > u_2 > 0$. If $3K^2 - s > 0$, then $\{S_j\} \in \mathfrak{T}$.

Proof:

Notice that when $3K^2 - s > 0$, then $f(u) := \zeta_j(s, u, K)/\xi_j(s, u, K)$ is increasing with $f(0) = 1$. Thus, $\zeta_j > \xi_j$ for all appropriate values of their arguments. We will now show that $\xi_j > \zeta_{j+1}$. The ratio ξ_j/ζ_{j+1} is increasing in j so it suffices to show that $\xi_1 > \zeta_2$. Note that when $8K^2 - s > 0$, then $g(u) := \xi_1(s, u, K)/\zeta_2(s, u, K)$ is decreasing with $g(K) = 1$, thus $\xi_1 > \zeta_2$. Having established the ordering $\zeta_j > \xi_j > \zeta_{j+1}$, the required result follows directly. $\square$

The restriction $3K^2 > s$ will not effect of general applicability of the algorithms that follow. As it will become apparent in the sequel, the user can easily specify an lower and upper bound on the possible range of values of K and s respectively so that it is guaranteed that $3K^2 > s$.

6.3.2 Construction of layered Brownian bridge

We return now to the setup we introduced at the beginning of Section 6.3. The distribution function of I can be written as

$$F(i) := \mathbb{P}[I \leq i] = \gamma \left(t, \frac{x-y}{2} + \frac{a_i - b_i}{2}, \frac{y-x}{2} + \frac{a_i - b_i}{2}, \frac{|y-x|}{2} + \frac{a_i + b_i}{2} \right) ,$$

for $i \geq 1$, with $F(0) = 0$. Notice that $\mathbb{P}[I = i] = \mathbb{P}[F(i-1) < U \leq F(i)]$, for a $U \sim \text{Un}(0, 1)$, therefore I can be simulated as follows. For any $i \geq 1$, let $\{S_j^i\} \in \mathfrak{T}$ be the alternating sequence converging to $F(i)$, as obtained from the representation of γ in (6.4). We also define the sequence $\{S_j^0\}$ with $S_j^0 = 0$ for all $j \geq 1$. Then, we set $I = i$ when $S_j^{i-1} < U < S_{j+1}^i$, for some odd j .

Having simulated the layer I of the Brownian bridge, we proceed to Step 4 of EA, that is to simulate the resulted layered Brownian bridge at the finite collection of time instances determined by the Poisson process. The most obvious approach for doing so is to use rejection sampling with candidates from $\mathbb{B}(t; x, y)$: for each proposed path we simulate its layer and if it differs from I the path is rejected. However this solution is theoretically and practically unappealing since if N_0 is the number of tries until (and including) the first success then:

$$\mathbb{E}[N_0] = \mathbb{E}[\mathbb{E}[N_0 | I]] = \sum_{i=1}^{\infty} \frac{1}{\mathbb{P}[I=i]} \times \mathbb{P}[I=i] = \infty .$$

Therefore we adopt an alternative rejection sampling procedure which proposes sample paths which are more likely to be accepted.

For the proposed construction it is important that we consider symmetric lay-

ers, i.e. we choose $a_i = b_i$, $i \geq 1$. Let $\mathbb{B}_{D_I}$ denote the law of the target layered Brownian bridge. We will propose from a mixture of two conditioned Brownian bridges. Specifically, for any $i \geq 1$ consider the events

$$\overline{\mathbf{M}}_i = \left\{ \sup_{0 \leq s \leq t} X_s \in [\bar{y} + a_{i-1}, \bar{y} + a_i] \right\}, \quad \underline{\mathbf{M}}_i = \left\{ \inf_{0 \leq s \leq t} X_s \in (\bar{x} - a_i, \bar{x} - a_{i-1}] \right\}.$$

Let $\mathbb{B}_{\overline{\mathbf{M}}_I}$ and $\mathbb{B}_{\underline{\mathbf{M}}_I}$ be the Brownian bridge law $\mathbb{B}(t; x, y)$ restricted to $\overline{\mathbf{M}}_I$ and $\underline{\mathbf{M}}_I$ respectively. Our rejection sampling algorithm will use candidate paths from

$$\mathbb{P}_{D_I} := \frac{1}{2} \mathbb{B}_{\overline{\mathbf{M}}_I} + \frac{1}{2} \mathbb{B}_{\underline{\mathbf{M}}_I}.$$

The following theorem, which can be easily proved, derives the appropriate acceptance probability. For a general subset of paths, say Δ , $\mathbb{B}[\Delta]$ is the probability under $\mathbb{B}(t; x, y)$ that a path is in Δ .

Theorem 6.4 $\mathbb{B}_{D_I} \ll \mathbb{P}_{D_I}$ with

$$\frac{d\mathbb{B}_{D_I}}{d\mathbb{P}_{D_I}}(X) = \frac{2 \mathbb{B}[\overline{\mathbf{M}}_I]}{\mathbb{B}[D_I]} \times \frac{\mathbb{I}[X \in D_I]}{1 + \mathbb{I}[X \in U_I \cap L_I]}.$$

Assume that N is the number of proposed paths until the first accepted one. Then,

$$\mathbb{E}[N | I] = \frac{2 \mathbb{B}[\overline{\mathbf{M}}_I]}{\mathbb{B}[D_I]} \leq \frac{2 \mathbb{B}[\overline{\mathbf{M}}_I]}{\mathbb{B}[U_I]} \leq \frac{2}{\mathbb{B}[\bar{x} - a_1 < X_s < \bar{y} + a_1, 0 \leq s \leq t]},$$

so, even for moderate a_1 , $\mathbb{E}[N]$ will be small.

Our algorithm w.p. 1/2 proposes a path $X \sim \mathbb{B}_{\overline{\mathbf{M}}_I}$ and with equal probability a path $X \sim \mathbb{B}_{\underline{\mathbf{M}}_I}$. Simulation from $\mathbb{B}_{\underline{\mathbf{M}}_I}$ (and by symmetry from $\mathbb{B}_{\overline{\mathbf{M}}_I}$) can be easily carried out following the decomposition of a Brownian bridge at its minimum in terms of two Bessel bridges summarized in Section 6.2.2. The calculation of the probability of accepting X requires also the simulation of the indicators $\mathbb{I}[X \in D_I]$ and $\mathbb{I}[X \in U_I \cap D_I]$ which will be carried out using the results about the hitting

times of a Bessel bridge given in Section 6.3.1. Analytically, we proceed as follows.

First, we simulate the minimum m of the proposed Brownian bridge X conditionally on that being in $(\bar{x} - a_I, \bar{x} - a_{I-1}]$, and then the time τ when the minimum is attained. Recall that the target layered path, thus the proposed path, needs to be simulated only at the time instances of the Poisson process $\Phi_{r(X)}$. In the simple case when $\Phi_{r(X)}$ has no points, the binary variable $\mathbb{I}[X \in D_I]$ is easy to simulate, since it is equal to 1 w.p.

$$\delta(\tau, 0, \bar{x} - m, \bar{y} + a_I - m) \times \delta(t - \tau, 0, \bar{y} - m, \bar{y} + a_I - m).$$

If $\mathbb{I}[X \in D_I] = 0$ the path X is rejected. In the opposite case, we proceed to the indicator $\mathbb{I}[X \in U_I \cap D_I]$. Conditionally on $\mathbb{I}[X \in D_I] = 1$, it is true that $\mathbb{I}[X \in U_I \cap L_I] = 0$, w.p.

$$\delta(\tau, 0, \bar{x} - m, \bar{y} + a_{I-1} - m; \bar{y} + a_I - m) \times \delta(t - \tau, 0, \bar{y} - m, \bar{y} + a_{I-1} - m; \bar{y} + a_I - m).$$

If $\mathbb{I}[X \in U_I \cap L_I] = 1$, then the proposed path is accepted w.p. 1/2, otherwise w.p.1.

In the general case when $\Phi_{r(X)}$ has a positive number of points, the procedure is similar. We simulate, m , τ , and the location of the proposed path at the time instances determined by $\Phi_{r(X)}$. Conditionally on this information, the path is retrieved as a product of Bessel bridges. Thus, the distribution of the binary variables $\mathbb{I}[X \in D_I]$ and $\mathbb{I}[X \in U_I \cap L_I]$ will be expressed in terms of products of hitting probabilities of the Bessel bridge and simulated following the alternating series method shown in Section 6.3.1.

6.3.3 Examples

The methodology we have developed applies easily to one-dimensional diffusions, V say, with positive and continuously differentiable coefficient $\sigma(\cdot)$. We first apply

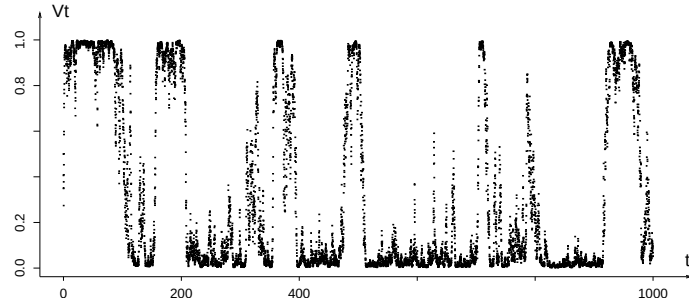


Figure 6.2: A skeleton of (6.5) generated by EA3. We used $\mu = 0.01$, $\nu = -0.02$, $\sigma = 1$ and starting point $V_0 = 0.5$. We applied iteratively EA3 for time increments $t = 0.1$ up to the time 1000. The increasing sequences $\{a_i\}$, $\{b_i\}$ involved in the algorithm were chosen as $a_i = b_i = 0.3 \cdot i$.

the standard transformation: $V_s \mapsto \int_{v^*}^{V_s} \sigma(z)^{-1} dz =: X_s$, for some arbitrary v^* in the state space of V , and then apply EA to X , provided that (C0)–(C3) hold for modified diffusion X .

As an illustration of EA3, we consider the following diffusion model arising in genetics (see e.g. ch.7 of Kloeden and Platen, 1992):

$$dV_s = (\mu + \nu V_s) ds + \sigma V_s(1 - V_s) dW_s, \quad (6.5)$$

with $\sigma > 0$ and μ, ν assumed to satisfy the restrictions $\mu > 0$ and $\mu + \nu < 0$ implying that V stays in $(0, 1)$. Under this parameter choice, straightforward calculations show that the corresponding unit coefficient process X , which moves in $\mathbf{R}$, satisfies conditions (C0)–(C3). Thus, we can apply EA3 to simulate sample paths of X and, equivalently, of V . Fig.6.2 shows a skeleton of V simulated using EA3.

6.4 Extensions

In this section we shall consider extensions of the results obtained and algorithms described thus far in the paper. We shall consider both multivariate and time-inhomogeneous generalizations.

6.4.1 Multivariate diffusions

We now consider unit diffusion coefficient d -dimensional diffusions satisfying

$$d\mathbf{X}_s = \boldsymbol{\alpha}(\mathbf{X}_s) ds + d\mathbf{W}_s, \quad s \in [0, t], \quad X_0 = \mathbf{x}. \quad (6.6)$$

Let $\mathbb{Q}$ and $\mathbb{W}$ denote the law of $\mathbf{X}$ and the d -dimensional Wiener measure respectively on $[0, t]$ for the initial condition $\mathbf{X}_0 = \mathbf{W}_0 = \mathbf{x}$. We will assume the following multi-dimensional equivalents to (C0)–(C3):

(MC0) The Radon-Nikodym derivative of $\mathbb{Q}$ w.r.t. $\mathbb{W}$ exists and it is given by the multivariate Girsanov's formula,

$$\frac{d\mathbb{Q}}{d\mathbb{W}} = \exp \left\{ \int_0^t \boldsymbol{\alpha}(\mathbf{X}_s) \cdot d\mathbf{X}_s - \int_0^t \frac{1}{2} \|\boldsymbol{\alpha}(\mathbf{X}_s)\|^2 ds \right\}.$$

(MC1.1) $\boldsymbol{\alpha}$ is continuously differentiable in all its arguments,

(MC1.2) there exists a function $A : \mathbf{R}^d \mapsto \mathbf{R}$ such that $\boldsymbol{\alpha}(\mathbf{u}) = \nabla A(\mathbf{u})$,

(MC2) there exists $l > -\infty$ such that $\phi(\mathbf{u}) := (\|\boldsymbol{\alpha}(\mathbf{u})\|^2 + \text{div } \boldsymbol{\alpha}(\mathbf{u}))/2 - l \geq 0$.

(MC3) the function $\exp\{-\|\mathbf{u} - \mathbf{x}\|^2/(2t) + A(\mathbf{u})\}$ is integrable in $\mathbf{u} \in \mathbf{R}^d$.

Notice that (MC1.2) above appears frequently in applications of multivariate diffusions. Analogously to the one-dimensional case, we consider the the epigraph of $s \mapsto \phi(\mathbf{X}_s)$,

$$\text{epi } \phi(\mathbf{X}) := \{ (s, u) \in [0, t] \times \mathbf{R}_+ : \phi(\mathbf{X}_s) \leq u \},$$

and the event

$$\Gamma := \{\Phi \subset \text{epi } \phi(\mathbf{X})\} = \bigcap_{j \geq 1} \{\phi(\mathbf{X}_{\psi_j}) < v_j\} .$$

Theorem 6.1 extends readily to the multivariate case, We define the probability measure $\mathbb{Z}$ w.r.t. the d -dimensional Wiener measure as

$$\frac{d\mathbb{Z}}{d\mathbb{W}} \propto \exp \{A(\mathbf{X}_t) - A(\mathbf{x})\} .$$

Theorem 6.5 (Multivariate Wiener-Poisson factorization) $\mathbb{Q}$ is the marginal law of $\mathbf{X}$ when $(\mathbf{X}, \Phi) \sim \mathbb{Z} \otimes \mathbb{L} | \Gamma$.

Theorem 6.5 leads directly to a simulation methodology as in the one-dimensional case. The most natural approach is to construct the proposed d -dimensional paths by means of d independent Brownian bridges with some corresponding layer choice. However, the assumption of unit diffusion coefficient is not as innocuous as in the one-dimensional case. Unfortunately, there is no obvious way to generalize this given the currently available methodology.

6.4.2 Time-inhomogeneous diffusions

EA can also be applied to diffusion processes of the form:

$$dX_s = \alpha(s, X_s) ds + dW_s, \quad s \in [0, t], \quad X_0 = x .$$

The necessary conditions for the applicability of EA are straightforward restatements of (C0)–(C3), thus we do not present them here explicitly. In the sequel, (IC0)–(IC3) will stand for the extension of (C0)–(C3) in the time-inhomogeneous context. In summary, we now ask that $\alpha \in C^{1,1}$, we define $A(s, u) = \int_0^u \alpha(s, z) dz$, and we let $\mathbb{Z}$ be the law determined through the following density:

$$\frac{d\mathbb{Z}}{d\mathbb{W}} \propto \exp \{A(t, X_t)\} .$$

Thus the Wiener-Poisson construction and its associated algorithms go through unchanged, setting

$$\phi(s, u) = \frac{1}{2} \left\{ \alpha^2(s, u) + \frac{\partial \alpha(s, u)}{\partial u} + 2 \frac{\partial A(s, u)}{\partial s} \right\} - l ,$$

where

$$l = \inf_{s \in [0, t]} \inf_{z \in \mathbf{R}} \left\{ \frac{1}{2} \alpha^2(s, z) + \frac{1}{2} \frac{\partial \alpha(s, z)}{\partial z} + \frac{\partial A(s, z)}{\partial s} \right\} .$$

Whereas (C2) is a standard assumption, its extension to this setting, i.e. the condition that $l > -\infty$ for l defined defined above, might not be easy to meet in great generality. A characteristic example is when X is Brownian motion with time-varying drift, that is $X_s = W_s + f(s)$ for some smooth function f .

Simulation of time-inhomogeneous diffusions is an essential ingredient of the exact simulation algorithm of one-sided boundary hitting times that we present in the next section.

6.5 Simulation of boundary hitting times

In this section we recast the methodology we have already developed to devise an algorithm for the simulation of diffusion hitting times of one-sided smooth boundaries. These problems can be equivalently stated as constant boundary crossing times for time-inhomogeneous diffusions, hence the extension of EA3 described in Section 6.4.2 will be required. We will restrict ourselves to the important special case of Brownian motion and consider the problem of the exact simulation of the first time that the Brownian motion exceeds a non-linear boundary $f \in C^2$. Thus, we will present a method for the simulation of the random variable

$$\zeta = \inf\{s \geq 0 : W_s = f(s)\} .$$

where W is a standard Brownian motion. Without loss of generality we will assume that $f(0) > 0$.

An attempt to simulate the time-inhomogeneous process $W_s - f(s)$ using EA with the objective to exploit the availability of results for crossing times of linear boundaries for the Brownian motion would fail since (IC2) does not hold. Notice that there are more general diffusion processes for which this inconvenience will not appear. We get around this difficulty by re-expressing the problem in terms of an Ornstein-Uhlenbeck process.

For an arbitrary tuning constant $a \in \mathbf{R}_+$, we consider the process

$$H_s := e^{-as/2} W_{(e^{as}-1)/a} ,$$

which satisfies $dH_s = -(a/2) H_s ds + dW_s$, $H_0 = 0$, for a suitable Brownian motion W . We define the modified boundary

$$g(s) = e^{-as/2} f((e^{as}-1)/a) ,$$

and consider the hitting time

$$\eta = \inf\{s \geq 0 : H_s = g(s)\} .$$

It is clear that $\zeta = (e^{a\eta}-1)/a$ a.s., thus it suffices to simulate η . We define the time-inhomogeneous diffusion $X_s := g(s) - H_s$ which satisfies

$$dX_s = (g'(s) + a g(s)/2 - a X_s/2) ds + dW_s, \quad X_0 = g(0) > 0 .$$

Now, η is the first passage time of X from 0.

We will simulate η through rejection sampling on the path space, using a variation of EA3 for time-inhomogeneous diffusions. The target diffusion will be

X defined above. It can be easily shown that conditions IC0–IC3 are satisfied. The algorithm will exploit the fact that the crossing times of linear boundaries of the proposed Brownian bridges are explicitly known and can be easily simulated.

We begin with the following well-known first passage time decomposition of $\mathbb{B}\mathbb{B}(t; x, y)$. Recall the definitions of m , τ and $\mathbb{S}$ given in Section 6.2.2 as the minimum of a path, the time that the minimum occurs, and the law of a Bessel bridge, respectively.

Theorem 6.6 *Assume that $x > 0$ and let η denote the first time that the bridge $X \sim \mathbb{B}\mathbb{B}(t; x, y)$ hits 0, where by convention we set $\eta = t$ if the hitting time does not occur in $[0, t]$.*

1. *For $\eta < t$, the law of $X \mid \eta$ admits the decomposition*

$$\mathbb{S}(\eta; x, 0) \otimes \mathbb{B}\mathbb{B}(\eta, t; 0, y) .$$

2. *For $\eta = t$, the law of $X - m \mid \eta, m, \tau$ can be decomposed as*

$$\mathbb{S}(\tau; x - m, 0) \otimes \mathbb{S}(\tau, t; 0, y - m) .$$

Therefore, the additional conditioning on the hitting time η has the effect that the proposed paths on $[0, t]$ are now expressed in terms of Bessel bridges. Although simulation of I is still possible (using the Bessel bridge hitting probabilities of Section 6.3.1), the algorithm of Section 6.3.2 for simulating the layered path cannot be applied, since that construction only applies to Brownian bridges.

We will nonetheless obtain lower and upper bounds for the proposed path by exploiting the well-known representation (see for instance Bertoin and Pitman, 1994) of the Bessel bridge starting from 0, in terms of three independent Brownian bridges. Analytically, for some $\delta > 0$ and a time increment $v > 0$, the stochastic

process

$$\sqrt{Z_{s,1}^2 + Z_{s,2}^2 + Z_{s,3}^2}, \quad s \in [0, v], \quad \text{where } Z_1 \sim \mathbb{B}\mathbb{B}(v; 0, \delta), \quad Z_2, Z_3 \sim \mathbb{B}\mathbb{B}(v; 0, 0),$$

is distributed according to $\mathbb{S}(v; 0, \delta)$. All Bessel bridges appearing in Theorem 6.6 are covered by this result due to the time-homogeneity of the Bessel bridge and its time-reversing property.

Our approach is to determine separate layers for each of the Brownian bridges that constitute the proposed Bessel bridge, using the techniques of Section 6.3. These layers will imply a bound on the range of the Bessel bridge. Simulation of the Bessel bridge conditional on this information will be done via the simulation of each constituent Brownian bridge conditional on its layer using the efficient rejection sampler of Section 6.3.2.

The precise mathematical construction is as follows. We will consider ten independent Brownian bridges. The first four

$$Z_1 \sim \mathbb{B}\mathbb{B}(\eta; x, 0), \quad Z_2, Z_3 \sim \mathbb{B}\mathbb{B}(\eta; 0, 0), \quad Z \sim \mathbb{B}\mathbb{B}(\eta, t; 0, y),$$

will be used in the case $\eta < t$, and the rest

$$Z_5 \sim \mathbb{B}\mathbb{B}(\tau; x - m, 0), \quad Z_6, Z_7 \sim \mathbb{B}\mathbb{B}(\tau; 0, 0),$$

$$Z_8 \sim \mathbb{B}\mathbb{B}(\tau, t; 0, y - m), \quad Z_9, Z_{10} \sim \mathbb{B}\mathbb{B}(\tau, t; 0, 0),$$

when $\eta = t$. For each bridge Z_i , we define $\bar{x}_i$ ($\bar{y}_i$) to be the minimum (maximum) between its starting and finishing point, and provide a positive increasing sequence $\{a_j\}$ for the development of symmetric layers around $\bar{x}_i$ and $\bar{y}_i$ in the way shown in Section 6.3. For simplicity, we assume the choice of an identical sequence $\{a_j\}$ for all ten bridges. In greater generality, one can select a different sequence for each bridge. Let I_i be the discrete random variable which specifies the layer of Z_i

and $\mathbf{I} = \{I_i; 1 \leq i \leq 10\}$. In this setting, $\mathcal{X}$ is defined on an enlargement of the canonical probability space for X , as

$$\mathcal{X} = (\eta, \mathbf{I}, m \mathbb{I}[\eta = t], \tau \mathbb{I}[\eta = t]).$$

Then, the proposed Brownian bridge and a bound on its range are constructed as follows.

- If $\eta < t$, then

$$X_s = \sqrt{Z_{s,1}^2 + Z_{s,2}^2 + Z_{s,3}^2}, \quad s \in [0, \eta], \quad X_s = Z_{s,4}, \quad s \in [\eta, t],$$

and $l(\mathbf{I}) \leq X_s \leq u(\mathbf{I})$ for all $s \in [0, t]$, where

$$l(\mathbf{I}) = \bar{x}_4 - a_{I_4}, \quad u(\mathbf{I}) = \max \left\{ \sqrt{(x + a_{I_1})^2 + a_{I_2}^2 + a_{I_3}^2}, \bar{y}_4 + a_{I_4} \right\}.$$

- If $\eta = t$, then

$$X_s = m + \sqrt{Z_{s,5}^2 + Z_{s,6}^2 + Z_{s,7}^2}, \quad s \in [0, \tau]; \quad X_s = m + \sqrt{Z_{s,8}^2 + Z_{s,9}^2 + Z_{s,10}^2}, \quad s \in [\tau, t],$$

and $l(\mathbf{I}) \leq X_s \leq u(\mathbf{I})$ for all $s \in [0, t]$, where $l(\mathbf{I}) = m$ and

$$u(\mathbf{I}) = m + \max \left\{ \sqrt{(x - m + a_{I_5})^2 + a_{I_6}^2 + a_{I_7}^2}, \sqrt{(y - m + a_{I_8})^2 + a_{I_9}^2 + a_{I_{10}}^2} \right\}.$$

Given $l(\mathbf{I})$, $u(\mathbf{I})$ we can compute the rate of the auxiliary Poisson process as

$$r(\mathcal{X}) = \sup \{ \phi(s, z) : s \in [0, t], z \in [l(\mathbf{I}), u(\mathbf{I})] \},$$

and, subsequently, complete the program prescribed in EA to decide whether or not to accept the proposed path and the corresponding hitting time. Notice, that in practice it is not necessary to find apriori ranges for all ten bridges since the

specific bridges required depend on the value of η .

Hitting times of two sided boundaries are in general not possible by this approach.

6.6 Discussion

In this paper we have reconstructed the law of a family of diffusions in terms of the Wiener and the Poisson measures. Based on this representation, we devised a simulation algorithm for un-solvable SDEs. Our algorithm, EA3, is described in Section 6.3 as an explicit finite construction for exact realisations from the finite dimensional distributions of the diffusion law. The methodology extends easily to time-inhomogeneous and multivariate diffusions, though the latter generalization is restricted to diffusions with gradient drift functions.

Crucial to our approach is the layered Brownian bridge construction which we initially describe in the EA3 context. In Section 6.5 we show that the same device used in a modified way can generate realisations of boundary crossing times for diffusions. The layered Brownian bridge construction is rather versatile. For instance it is straightforward to construct diffusion sample paths conditioned to remain within a region, E , (compact or otherwise) by just arranging that the initial layer ($I = 1$) corresponds to the sample path remaining within E . Furthermore diffusions with finite entrance boundaries such as

$$dX_t = -\tan(X_t) dt + dW_t ,$$

can be easily simulated by choosing sequences of layers with edges converging to $\pm\pi/2$.

Acknowledgements

The authors would like to thank Bruno Casella for several helpful discussions on the subjects of this paper. They also gratefully acknowledge support from EPSRC grant GR/S61577/01. One of the authors research was funded by the Greek State Scholarship's Foundation.

Bibliography

- Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, **9**, 385–426.
- (2002) Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, **70**, 223–262.
- (2005) Closed-form likelihood expansions for multivariate diffusions. Working paper.
- Amit, Y. (1991) On rates of convergence of stochastic relaxation for Gaussian and non-Gaussian distributions. *J. Multivariate Anal.*, **38**, 82–99.
- Andrews, D. W. K. (1987) Consistency in nonlinear econometric models: a generic uniform law of large numbers. *Econometrica*, **55**, 1465–1471.
- Arnold, B., Castillo, E. and Sarabia, J. (1999) *Conditional Specification of Statistical Models*. Springer.
- Arnold, B. C. and Strauss, D. (1988) Bivariate distributions with exponential conditionals. *J. Amer. Statist. Assoc.*, **83**, 522–527.
- Asmussen, S., Glynn, P. and Pitman, J. (1995) Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, **5**, 875–896.
- Athreya, K. B. and Ney, P. (1978) A new approach to the limit theory of recurrent Markov chains. *Trans. Amer. Math. Soc.*, **245**, 493–501.

- Attouch, H. (1984) *Variational convergence for functions and operators*. Applicable Mathematics Series. Boston, MA: Pitman (Advanced Publishing Program).
- Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **64**, 253–280.
- Beck, A. (1963) On the strong law of large numbers. In *Ergodic Theory (Proc. Internat. Sympos., Tulane Univ., New Orleans, La., 1961)*, 21–53. New York: Academic Press.
- Bertoin, J. and Pitman, J. (1994) Path transformations connecting Brownian bridge, excursion and meander. *Bull. Sci. Math.*, **118**, 147–166.
- Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2005a) A factorisation of diffusion measure and finite sample path constructions. Submitted.
- (2005b) Monte carlo maximum likelihood estimation for discretely observed diffusion processes. Submitted.
- (2005c) Retrospective exact simulation of diffusion sample paths with applications. Submitted.
- (2006a) Markov chain Monte Carlo for exact inference for discretely observed diffusions. In progress.
- Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006b) Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **68**, 1–29.
- Beskos, A. P. and Roberts, G. (2005) Exact simulation of diffusions. *Ann. Appl. Probab.*, **15**.
- Bibby, B. M., Jacobsen, M., and Sørensen, M. (2002) Estimating functions for discretely sampled diffusion-type models. In *Handbook of Financial Econometrics*

- (eds. Y. Aït-Sahalia and L. P. Hansen). Amsterdam, The Netherlands: North Holland.
- Bibby, B. M. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.
- (1996) On estimation for discretely observed diffusions: A review. *Theory Stoch. Process.*, **2**, 49–56.
- (1997) A hyperbolic diffusion model for stock prices. *Finance Stoch.*, **1**, 25–41.
- (2001) Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scand. J. Statist.*, **28**, 99–112.
- Billingsley, P. (1961) The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.*, **12**, 788–792.
- Black, F. and Scholes, M. S. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637–659.
- Borodin, A. N. and Salminen, P. (2002) *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Basel: Birkhäuser Verlag, second edn.
- Borovkov, A. A. and Foss, S. G. (1992) Stochastically recursive sequences and their generalizations. *Siberian Adv. Math.*, **2**, 16–81. Siberian Advances in Mathematics.
- Brandt, M. W. and Santa-Clara, P. (2002) Simulated likelihood estimation of multivariate diffusions with an application to interest rates and exchange rates with stochastic volatility. *Journal of Financial Economics*, **63**, 161–210.
- Breyer, L. and Roberts, G. (2000) Some multi-step coupling constructions for Markov chains. *Tech. rep.*, Lancaster University.

- (2001) Catalytic perfect simulation. *Methodology and Computing in Applied Probability*, **3**, 161–177.
- Brooks, S. P. and Roberts, G. O. (1998) Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, **8**, 319–335.
- Burke, G. (1965) A uniform ergodic theorem. *Ann. Math. Statist.*, **36**, 1853–1858.
- Casella, B. and Roberts, G. (2005) Exact monte carlo estimation of expectations of diffusions with barriers. In progress.
- Chan, K., Karolyi, A. G., Longstaff, F. A. and Sanders, A. B. (1992) An empirical comparison of alternative models of the short-term interest rate. *J. Finance*, **47**, 1209–1227.
- Chan, K. S. and Ledolter, J. (1995) Monte Carlo EM estimation for time series models involving counts. *J. Amer. Statist. Assoc.*, **90**, 242–252.
- Choirat, C., Hess, C. and Seri, R. (2003) A functional version of the Birkhoff ergodic theorem for a normal integrand: a variational approach. *Ann. Probab.*, **31**, 63–92.
- Corcoran, J. N. and Tweedie, R. L. (2002) Perfect sampling from independent Metropolis-Hastings chains. *J. Statist. Plann. Inference*, **104**, 297–314.
- Cowles, M. K. and Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.*, **91**, 883–904.
- Cox, J. C., Ingersoll, Jr., J. E. and Ross, S. A. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.
- Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **39**, 1–38. With discussion.
- Devroye, L. (1986) *Nonuniform random variate generation*. New York: Springer-Verlag.
- Diaconis, P. and Stroock, D. (1991) Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, **1**, 36–61.
- Djuric, P. M. and Kotecha, J. H. (1999) Gibbs sampling approach for generation of truncated multivariate gaussian random variables. *Proceedings of IEEE ICASSP-99*, **3**, 1757–1760.
- Doob, J. L. (1949) Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, **20**, 393–403.
- Dudley, R. M. (1999) *Uniform central limit theorems*, vol. 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge: Cambridge University Press.
- Duffee, G. and Stanton, R. (2004) Estimation of dynamic term structure models. Working paper.
- Durham, G. B. and Gallant, A. R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *J. Bus. Econom. Statist.*, **20**, 297–338. With comments and a reply by the authors.
- Elerian, O. (1999) *Simulation estimation of continuous time series models with applications to finance*. Ph.D. thesis, Nuffield College, Oxford.
- Elerian, O., Chib, S. and Shephard, N. (2001) Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, **69**, 959–993.
- Eraker, B. (2001) MCMC analysis of diffusion models with application to finance. *J. Bus. Econom. Statist.*, **19**, 177–191.

- Fearnhead, P., Papaspiliopoulos, O. and Roberts, G. (2006) Exact particle filters for partially observed diffusions. In progress.
- Fill, J. A. (1998) An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.*, **8**, 131–162.
- Fill, P. and Machida, D. (1999) Realisable and stochastic monotonicity. *SIAM Rev.*, **41**, 45–76 (electronic).
- Forsythe, G. E. (1972) von Neumann's comparison method for random sampling from the normal and other distributions. *Math. Comp.*, **26**, 817–826.
- Fort, G. and Moulines, E. (2003) Convergence of the Monte Carlo expectation maximization for curved exponential families. *Ann. Statist.*, **31**, 1220–1259.
- Foss, S. G. and Tweedie, R. L. (1998) Perfect simulation and backward coupling. *Comm. Statist. Stochastic Models*, **14**, 187–203. Special issue in honor of Marcel F. Neuts.
- Gallant, A. R. and Long, J. R. (1997) Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika*, **84**, 125–141.
- Gallant, A. R. and Tauchen, G. (1996) Which moments to match? *Econometric Theory*, **12**, 657–681.
- Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, **85**, 398–409.
- Geweke, J. (1991) Efficient simulation from the multivariate normal and Student- t distributions subject to linear constraints. In *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface* (ed. E. M. Keramidas), 571–578. Interface Foundation of North America.
- Geyer, C. J. (1994) On the convergence of Monte Carlo maximum likelihood calculations. *J. Roy. Statist. Soc. Ser. B*, **56**, 261–274.

- Geyer, C. J. and Thompson, E. A. (1992) Constrained Monte Carlo maximum likelihood for dependent data. *J. Roy. Statist. Soc. Ser. B*, **54**, 657–699. With discussion and a reply by the authors.
- Giesy, D. P. (1976) Strong laws of large numbers for independent sequences of Banach space-valued random variables. In *Probability in Banach spaces (Proc. First Internat. Conf., Oberwolfach, 1975)*, 89–99. Lecture Notes in Math., Vol. 526. Berlin: Springer.
- Gillespie, D. T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Computational Phys.*, **22**, 403–434.
- (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Gobet, E. (2002) LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, **38**, 711–737.
- Godambe, V. P. and Heyde, C. C. (1987) Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.*, **55**, 231–244.
- Goel, N. S. and Richter-Dyn, N. (1974) *Stochastic models in biology*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London.
- Gourieroux, C. and Jasiak, J. (2003) Jacobi process and smooth transitions. Working paper.
- Gourieroux, C., Monfort, A. and Renault, E. (1993) Indirect inference. *J. Applied Econometrics*, **8**, 85–118.
- Green, P. J. and Murdoch, D. J. (1999) Exact sampling for Bayesian inference:

- towards general purpose algorithms. In *Bayesian statistics, 6 (Alcoceber, 1998)*, 301–321. New York: Oxford Univ. Press.
- Häggström, O., van Lieshout, M.-C. N. M. and Møller, J. (1999) Characterization results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli*, **5**, 641–658.
- Häggström, O. and Nelander, K. (1998) Exact sampling from anti-monotone systems. *Statist. Neerlandica*, **52**, 360–380.
- Hajivassiliou, V., McFadden, D. and Ruud, P. (1996) Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics*, **72**, 85–134.
- Hajivassiliou, V. A. and McFadden, D. L. (1998) The method of simulated scores for the estimation of LDV models. *Econometrica*, **66**, 863–896.
- Heyde, C. C. (1997) *Quasi-likelihood and its application*. Springer Series in Statistics. New York: Springer-Verlag.
- Jamshidian, M. and Jennrich, R. I. (2000) Standard errors for EM estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **62**, 257–270.
- Jensen, B. and Poulsen, R. (2002) Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives*, **9**, 18–32.
- Jerrum, M. and Sinclair, A. (1989) Approximating the permanent. *SIAM J. Comput.*, **18**, 1149–1178.
- Johnson, V. E. (1996) Studying convergence of Markov chain Monte Carlo algorithms using coupled sample paths. *J. Amer. Statist. Assoc.*, **91**, 154–166.
- Jones, C. (1998) A simple bayesian method for the analysis of diffusion and jump-diffusion processes. Working paper.

- Karatzas, I. and Shreve, S. E. (1991) *Brownian motion and stochastic calculus*, vol. 113 of *Graduate Texts in Mathematics*. New York: Springer-Verlag, second edn.
- Karlin, S. (1968) *Total positivity. Vol. I*. Stanford, Calif: Stanford University Press.
- Karlin, S. and Taylor, H. M. (1981) *A second course in stochastic processes*. New York: Academic Press Inc. [Harcourt Brace Jovanovich Publishers].
- Kendall, M. and Stuart, A. (1977) *The Advanced Theory of Statistics. Vol. I: Distribution Theory (4th Ed)*. Charles Griffin & Co.
- Kendall, W. (1997) On some weighted Boolean models. In *Advances in Theory and Applications of Random Sets* (ed. D. Jeulin), 105–120. World Scientific Publishing Company.
- Kendall, W. S. (1998) Perfect simulation for the area-interaction point process. In *Probability towards 2000 (New York, 1995)*, vol. 128 of *Lecture Notes in Statist.*, 218–234. New York: Springer.
- (2004) Geometric ergodicity and perfect simulation. *Electron. Comm. Probab.*, **9**, 140–151 (electronic).
- Kendall, W. S. and Møller, J. (2000) Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Adv. in Appl. Probab.*, **32**, 844–865.
- Kendall, W. S. and Thönnies, E. (1999) Perfect simulation in stochastic geometry. *Pattern Recognition*, **32**, 1569–1586.
- Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.

- (2000) Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.*, **27**, 65–82.
- Kessler, M. and Paredes, S. (2002) Computational aspects related to martingale estimating functions for a discretely observed diffusion. *Scand. J. Statist.*, **29**, 425–440.
- Kessler, M. and Sørensen, M. (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**, 299–314.
- Kimura, M. and Ohta, T. (1971) *Theoretical aspects of population genetics*. Princeton University Press.
- Kingman, J. F. C. (1993) *Poisson processes*, vol. 3 of *Oxford Studies in Probability*. New York: The Clarendon Press Oxford University Press. Oxford Science Publications.
- Kloeden, P. E. and Platen, E. (1992) *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*. Berlin: Springer-Verlag.
- Lerche, H. R. (1986) *Boundary crossing of Brownian motion*, vol. 40 of *Lecture Notes in Statistics*. Berlin: Springer-Verlag. Its relation to the law of the iterated logarithm and to sequential analysis.
- Lindvall, T. (1992) *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: John Wiley & Sons Inc. A Wiley-Interscience Publication.
- Louis, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, **44**, 226–233.
- Maruyama, G. (1953) Markov processes and stochastic equations. *Nat. Sci. Rep. Ochanomizu Univ.*, **4**, 40–43.

- (1955) Continuous Markov processes and stochastic equations. *Rend. Circ. Mat. Palermo (2)*, **4**, 48–90.
- McAdams, H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, **94**, 814–819.
- McKean, H. (1969) *Stochastic Integrals*. Academic Press.
- Meng, X.-L. and Rubin, D. B. (1991) Using EM to obtain asymptotic variance-covariance matrices: The SEM algorithm. *Journal of the American Statistical Association*, **86**, 899–909.
- Merton, R. C. (1971) Optimum consumption and portfolio rules in a continuous-time model. *J. Econom. Theory*, **3**, 373–413.
- Meyn, S. P. and Tweedie, R. L. (1993) *Markov chains and stochastic stability*. Communications and Control Engineering Series. London: Springer-Verlag London Ltd.
- (1994) Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, **4**, 981–1011.
- Milstein, G. N. (1974) Approximate integration of stochastic differential equations. *Theory Probab. Appl.*, **19**, 557–562.
- Mira, A., Møller, J. and Roberts, G. O. (2001) Perfect slice samplers. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **63**, 593–606.
- Møller, J. (1999) Perfect simulation of conditionally specified models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **61**, 251–264.
- Montana, G. (2002) *Small sets and domination in perfect simulation*. Ph.D. thesis, University of Warwick.

- Mourier, E. (1953) Eléments aléatoires dans un espace de Banach. *Ann. Inst. H. Poincaré*, **13**, 161–244.
- Murdoch, D. and Green, P. (1998a) Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, **25**, 483–502.
- Murdoch, D. J. and Green, P. J. (1998b) Exact sampling from a continuous state space. *Scand. J. Statist.*, **25**, 483–502.
- Nummelin, E. (1978) A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, **43**, 309–318.
- Obuhov, A. M. (1959) Description of turbulence in terms of Lagrangian variables. In *Advances in Geophysics, Vol. 6 (Symposium on Atmospheric Diffusion and Air Pollution, Oxford, 1958)*, 113–116. New York: Academic Press.
- Oksendal, B. K. (1998) *Stochastic Differential Equations: An Introduction With Applications*. Springer-Verlag.
- Papaspiliopoulos, O. and Roberts, G. (2004) Retrospective MCMC methods for Dirichlet process hierarchical models. Submitted.
- Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003) Non-centered parameterizations for hierarchical models and data augmentation. In *Bayesian statistics, 7 (Tenerife, 2002)*, 307–326. New York: Oxford Univ. Press. With a discussion by Alan E. Gelfand, Ole F. Christensen and Darren J. Wilkinson, and a reply by the authors.
- Pardoux, É. and Pignol, M. (1984) Étude de la stabilité de la solution d’une EDS bilinéaire à coefficients périodiques. Application au mouvement des pales d’hélicoptère. In *Analysis and optimization of systems, Part 2 (Nice, 1984)*, vol. 63 of *Lecture Notes in Control and Inform. Sci.*, 92–103. Berlin: Springer.

- Pedersen, A. R. (1995a) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli*, **1**, 257–279.
- (1995b) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, **22**, 55–71.
- Peskir, G. (2000) *From uniform laws of large numbers to uniform ergodic theorems*, vol. 66 of *Lecture Notes Series (Aarhus)*. University of Aarhus, Department of Mathematics, Aarhus.
- Philippe, A. and Robert, C. P. (2003) Perfect simulation of positive gaussian distributions. *Statistics and Computing*, **13**, 179–186.
- Pötzelberger, K. and Wang, L. (2001) Boundary crossing probability for Brownian motion. *J. Appl. Probab.*, **38**, 152–164.
- Prakasa Rao, B. L. S. (1999) *Statistical inference for diffusion type processes*, vol. 8 of *Kendall's Library of Statistics*. London: Edward Arnold.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992) *Numerical recipes in C*. Cambridge: Cambridge University Press, second edn. The art of scientific computing.
- Propp, J. and Wilson, D. (1996) Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, **9**, 223–252.
- (1998a) Coupling from the past: a user's guide. In *Microsurveys in discrete probability (Princeton, NJ, 1997)*, vol. 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, 181–192. Providence, RI: Amer. Math. Soc.
- Propp, J. G. and Wilson, D. B. (1998b) How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed

- graph. *J. Algorithms*, **27**, 170–217. 7th Annual ACM-SIAM Symposium on Discrete Algorithms (Atlanta, GA, 1996).
- Revuz, D. and Yor, M. (1991) *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Berlin: Springer-Verlag.
- Rimmer, D., Doucet, A. and Fitzgerald, W. (2005) Particle filters for stochastic differential equations of non-linear diffusions. Technical report CUED/F-INFENG/TR.497.
- Robert, C. (1995) Simulation of truncated normal variables. *Statistics and Computing*, **5**, 121–125.
- Roberts, G. and Rosenthal, J. (2002) One-Shot coupling for certain stochastic recursive sequences. *Stochastic Processes and their Applications*, **99**, 195–208.
- Roberts, G. and Tweedie, R. (2006) Understanding MCMC. To be published.
- Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P. (2004) Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66**, 369–393.
- Roberts, G. O. and Rosenthal, J. S. (1996) Quantitative bounds for convergence rates of continuous time Markov processes. *Electron. J. Probab.*, **1**, no. 9, approx. 21 pp. (electronic).
- Roberts, G. O. and Stramer, O. (2001) On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**, 603–621.
- Rodriguez-Yam, G. A., Davis, R. A. and Scharf, L. L. (2002) A bayesian model and Gibbs sampler for hyperspectral imaging. *Proceedings 2002 IEEE Sensor Array and Multichannel Signal Processing Workshop*, 105–109.

- Rogers, L. C. G. and Williams, D. (2000) *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge: Cambridge University Press. Ito calculus, Reprint of the second (1994) edition.
- Rosenthal, J. S. (1995) Minorization conditions and convergence rates for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.*, **90**, 558–566.
- Santa-Clara, P. (1995) Simulated likelihood estimation of diffusions with an application to the short-term interest rate. Working paper.
- Shepp, L. A. (1979) The joint density of the maximum and its location for a Wiener process with drift. *J. Appl. Probab.*, **16**, 423–427.
- Sherman, R. P., Ho, Y.-Y. K. and Dalal, S. R. (1999) Conditions for convergence of Monte Carlo EM sequences with an application to product diffusion modeling. *Econom. J.*, **2**, 248–267.
- Shiga, T. (1985) Mathematical results on the stepping stone model in population genetics. In *Population genetics and molecular evolution (Mishima, 1984)*, 267–279. Tokyo: Japan Sci. Soc. Press.
- Smith, A. A. (1993) Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics*, **8S**, 63–84.
- Sørensen, H. (2004) Parametric inference for diffusion processes observed at discrete points in time: a survey. *Internat. Statist. Rev.*, **72**, 337–354.
- Sørensen, M. (1997) Estimating functions for discretely observed diffusions: a review. In *Selected Proceedings of the Symposium on Estimating Functions (Athens, GA, 1996)*, vol. 32 of *IMS Lecture Notes Monogr. Ser.*, 305–325. Hayward, CA: Inst. Math. Statist.
- (1999) On asymptotics of estimating functions. *Braz. J. Probab. Stat.*, **13**, 111–136.

- Talagrand, M. (1987) The Glivenko-Cantelli problem. *Ann. Probab.*, **15**, 837–870.
- Thorisson, H. (2000) *Coupling, Stationarity and Renegeration*. Springer.
- Tierney, L. (1994) Markov chains for exploring posterior distributions. *Ann. Statist.*, **22**, 1701–1762. With discussion and a rejoinder by the author.
- Troughton, P. T. and Godsill, S. (1999) MCMC methods for restoration of quantised time series. *Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*, **2**, 447–451.
- Tucker, H. G. (1959) A generalization of the Glivenko-Cantelli theorem. *Ann. Math. Statist.*, **30**, 828–830.
- Varga, R. S. (1962) *Matrix Iterative Analysis*. Prentice-Hall Inc.
- Wagner, W. (1988a) Monte Carlo evaluation of functionals of solutions of stochastic differential equations. Variance reduction and numerical examples. *Stochastic Anal. Appl.*, **6**, 447–468.
- (1988b) Unbiased multi-step estimators for the Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.*, **79**, 336–352.
- (1989) Unbiased Monte Carlo estimators for functionals of weak solutions of stochastic differential equations. *Stochastics Stochastics Rep.*, **28**, 1–20.
- Wagner, W. and Platen, E. (1978) Approximation of itô integral equations. Akad. der Wiss. der DDR, Berlin.
- Wald, A. (1949) Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statistics*, **20**, 595–601.
- Wei, G. C. G. and Tanner, M. A. (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

- Wilson, D. (2000a) Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP). In *Monte Carlo Methods* (ed. N. Madras), vol. 26 of *Fields Institute Communications*, 141–176. American Mathematical Society.
- Wilson, D. B. (2000b) How to couple from the past using a read-once source of randomness. *Random Structures Algorithms*, **16**, 85–113.
- (2000c) Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP). In *Monte Carlo methods (Toronto, ON, 1998)*, vol. 26 of *Fields Inst. Commun.*, 143–179. Providence, RI: Amer. Math. Soc.
- Zhou, H. (2001) Finite sample properties of EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model. *Journal of Computational Finance*, **5**, 89–122.