Forward and reverse representations for Markov chains*

G. N. Milstein[b], J.G.M. Schoenmakers[a], V. Spokoiny[a] †

February 8, 2007

Keywords: transition density estimation, forward and reverse Markov chains, Monte Carlo simulation, estimation of risk

AMS 2000 Subject Classification: 60J05, 60H10, 62G07, 65C05.

*This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 Economic Risk, and the DFG Research Center MATHEON in Berlin.

†a) Weierstrass Institut für Angewandte Analysis und Stochastik, Berlin, Germany, {schoenma,spokoiny}@wias-berlin.de b) Ural State University, Ekaterinburg, Russia, grigori.milstein@usu.ru

Abstract

In this paper we carry over the concept of reverse probabilistic representations developed in Milstein, Schoenmakers, Spokoiny (2004) for diffusion processes, to discrete time Markov chains. We outline the construction of reverse chains in several situations and apply this to processes which are connected with jump-diffusion models and finite state Markov chains. By combining forward and reverse representations we then construct transition density estimators for chains which have root-N accuracy in any dimension and consider some applications.

1 Introduction

The paper is concerned with the problem of obtaining the transition density for a Markov chain described by its one-step transition probabilities. In general, we do not assume that the underlying Markov chain is autonomous, although this case is discussed as a particular example. The problem of calculating the transition density naturally arise in different areas such as risk analysis and environmental modelling. Typically there is no analytic closed form solution for the target density. Instead, a Monte Carlo method obtained by combining probabilistic representations with statistical (e.g. kernel) density estimation is presented. Classical density estimators for random processes such as the Parzen-Rosenblatt estimator (see for instance Silverman (1986)) suffer typically from "curse of dimensionality" which means a poor quality of estimation when the dimension of the state space is large. So, the computational demand for achieving a desirable estimation accuracy is enormous and often unfeasible.

In a diffusion setting, Milstein, Schoenmakers, Spokoiny (2004) have developed a new approach which is based on forward and reverse simulation of the underlying diffusion process. It is shown that this forward-reverse density estimator (FRE) is basically root-N consistent in any dimension, this in contrast to the Parzen-Rosenblatt estimator which has accuracy $N^{-1/(4+d)}$ for dimension d. The forward-reverse estimator has turned out to be very useful for many practical applications. For example, in van den Berg, Heemink, Lin, Schoenmakers (2003), Spivakovskaya, Heemink, Milstein, Schoenmakers (2005), and Spivakovskaya, Heemink, Schoenmakers (2006), the FRE is applied successfully to the estimation of pollutant concentrations in small coastal water regions, which are caused by a certain calamity at another place. In the latter applications the pollutants are modelled by a diffusion process. In other areas however, for instance finance, the relevant underlying quantities are often modelled by time series, hence *discrete-time* Markov processes. While the analytical tools for diffusion theory are essentially connected with parabolic partial differential equations, the analytical formalism for discrete-time Markov processes is connected with integral equations. So for such processes one can not relay on diffusion theory and thus a discrete time version of the theory of forward reverse estimation in terms of integral equations is called for. This theory is provided in the present article.

For convenience of the reader we summarize the main results of Milstein, Schoenmakers, Spokoiny (2004) in Section 2. The rest of the paper is organized as follows. In Section 3 we introduce general forward representations for discrete time Markov chains. From the forward chains a family of reverse chains are derived in the spirit of Milstein, Schoenmakers, Spokoiny (2004) in Section 4. Section 5 formalizes general variance reduction for both forward and reverse chains. In Section 6 reverse chains are derived for a special class of autonomous discrete-time Markov chains and some examples are given. In Section 7 we give an application to jump-diffusion models and in Section 8 we deal with Markov processes driven by finite state Markov chains. The forward-reverse density estimator for Markov chains is discussed finally in Section 9.

2 Density estimation for diffusions based on forward reverse representations

Consider a stochastic differential equation (SDE) in the Itô sense

$$dX = a(s,X)ds + \sigma(s,X)dW(s), \qquad t_0 \le t \le s \le T, \qquad X(t) = x, \tag{1}$$

where $X = (X^1, ..., X^d)^\top$, $a = (a^1, ..., a^d)^\top$ are d-dimensional vectors, $W = (W^1, ..., W^m)^\top$ is an m-dimensional standard Wiener process, and $\sigma = \{\sigma^{ij}\}$ is a $d \times m$-matrix, $m \ge d$. It is assumed that the $d \times d$-matrix $b := \sigma\sigma^\top$, $b = \{b^{ij}\}$, is of full rank for every (s,x), $s \in [t_0, T]$, $x \in R^d$. The functions $a^i(s,x)$ and $\sigma^{ij}(s,x)$ are assumed to be bounded and to have bounded derivatives of any order, which implies existence and uniqueness of the solution $X^{t,x}(s)$, $X^{t,x}(t) = x$, $t_0 \le t \le s \le T$, of (1), smoothness of the transition density $p(t,x,s,y)$ of the Markov process X, and existence of all the moments of $p(\cdot,\cdot,\cdot,y)$.
The solution of SDE (1) may be approximated by different numerical methods, see Kloeden and Platen (1992), Milstein and Tretyakov (2004).

2.1 The Parzen-Rosenblatt forward estimator (FE)

Let $\bar{X}^{t,x}$ be a numerical approximation of the process $X^{t,x}$ and let $\bar{X}_n^{t,x}(T)$, $n = 1, \ldots, N$, be a sample of independent realizations of $\bar{X}^{t,x}(T)$. Then one may estimate the transition density $p(t,x,T,y)$ from this sample by using standard techniques of non-parametric statistics such as the classical kernel (Parzen-Rosenblatt) estimator. The kernel (Parzen-Rosenblatt) forward density estimator with a kernel K and a bandwidth δ is given by

$$\hat{p}_{FE}(t,x,T,y) = \frac{1}{N\delta^d} \sum_{n=1}^{N} K\left(\frac{\bar{X}_n^{t,x}(T) - y}{\delta}\right), \tag{2}$$

see Devroye and Györfi (1985), Silverman (1986). For example, in (2) one could take the Gaussian kernel $K(x) = (2\pi)^{-d/2}\exp(-|x|^2/2)$. Here δ should decrease to zero as N increases while $N\delta^d \to \infty$. It is well known that the quality of density estimation strongly depends on the bandwidth δ and the choice of a suitable bandwidth is a delicate issue (see e.g. Devroye and Györfi (1985)). Even an optimal choice of the bandwidth δ leads to quite poor estimation quality, in particular for large dimension d. More specifically, if the underlying density is known to be two times continuously differentiable then the optimal bandwidth δ is of order $N^{-1/(4+d)}$ leading to the accuracy of order $N^{-2/(4+d)}$, see Scott (1992) or Silverman (1986). For larger d, this would require a huge sample size N for providing a reasonable accuracy of estimation. In the statistical literature this problem is referred to as "curse of dimensionality".

2.2 The reverse estimator (RE)

In order to proceed with more sophisticated density estimators we introduce a reverse diffusion system for (1). We first introduce a reversed time variable $\tilde{s} = T + t - s$ and define

$$\begin{aligned} \tilde{a}(\tilde{s}, y) &= a(T + t - \tilde{s}, y), \\ \tilde{b}(\tilde{s}, y) &= b(T + t - \tilde{s}, y), \\ \tilde{\sigma}(\tilde{s}, y) &= \sigma(T + t - \tilde{s}, y). \end{aligned}$$

Then we introduce a vector process $Y^{t,y,1} \in \mathbb{R}^d$ and a scalar process $\mathcal{Y}^{t,y}$ governed by the reverse time stochastic system

$$\begin{array}{ll} dY = \alpha(s, Y)ds + \tilde{\sigma}(s, Y)d\tilde{W}(s), & Y(t) = y, \\ d\mathcal{Y} = c(s, Y)\mathcal{Y}ds, & \mathcal{Y}(t) = 1, \quad t_0 \le t \le s \le T, \end{array} \tag{3}$$

with $\tilde{W}$ being an m-dimensional standard Wiener process and

$$\begin{aligned} \alpha^i &= \sum_{j=1}^{d} \frac{\partial \tilde{b}^{ij}}{\partial y^j} - \tilde{a}^i, \\ c &= \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2 \tilde{b}^{ij}}{\partial y^i \partial y^j} - \sum_{i=1}^{d} \frac{\partial \tilde{a}^i}{\partial y^i}. \end{aligned}$$

It is possible to construct an alternative density estimator in terms of the reverse system (3). Suppose that $(\bar{Y}_m^{t,y}, \bar{\mathcal{Y}}_m^{t,y,1})$, $m = 1, \ldots, M$, is an i.i.d. sample of numerical solutions of (3). Then a pure reverse estimator is given by

$$\hat{p}_{RE}(t, x, T, y) := \frac{1}{M\delta^d} \sum_{m=1}^{M} K\left(\frac{x - \bar{Y}_m^{t,y}(T)}{\delta}\right) \bar{\mathcal{Y}}_m^{t,y,1}(T). \tag{4}$$

In fact, the reverse estimator (4) can be obtained as a side case from the forward-reverse estimator presented below.

2.3 The forward-reverse estimator (FRE)

By combining the forward (1) and reverse (3) estimators via the Chapman - Kolmogorov equation with respect to an intermediate time t^*, one can construct the forward-reverse estimator (see Milstein, Schoenmakers, Spokoiny (2004)),

$$\hat{p}_{FRE}(t,x,T,y) = \frac{1}{MN\delta^d} \sum_{m=1}^{M} \sum_{n=1}^{N} K\left(\frac{\bar{X}_n^{t,x}(t^*) - \bar{Y}_m^{t^*,y}(T)}{\delta}\right) \bar{\mathcal{Y}}_m^{t^*,y,1}(T). \quad (5)$$

It is shown that the forward-reverse estimator (5) has superior properties in comparison with density estimators based on pure forward (2) or pure reverse (4) representations. Obviously, by taking $t^* = T$ and $t^* = 0$, the estimator (5) collapses to the pure forward estimator (2) and pure reverse estimator (4), respectively.

For estimating a target value p by an estimator $\hat{p}$ we define the accuracy of the estimator by

$$\text{Accuracy}(\hat{p}) := \epsilon(\hat{p}) := \sqrt{E(\hat{p}-p)^2} = \sqrt{\text{Var}(\hat{p}) + \text{Bias}^2(\hat{p})}. \quad (6)$$

Loosely speaking, for a first order kernel applied in (5), any choice of $0 < t^* < T$, and a bandwidth choice $\delta_N = \mathcal{O}(N^{-1/(4+d)})$, the FRE has root-N ($\mathcal{O}(N^{-1/2})$) accuracy for dimension $d \leq 4$. For $d > 4$ root-N accuracy is lost but then the FRE accuracy order is still the square of the FE/RE accuracy order (see Table 1). Moreover, it can be shown that root-N accuracy of (5) can also be achieved for $d > 4$ by using higher order kernels in (5).

By definition (6) it is possible to relate the "expected" accuracy of the different density estimators to the number of simulated trajectories involved. However, simulating trajectories is not the only costly issue in the density estimation. For all estimators one has to evaluate a functional of the simulated trajectories. In case of the FE and RE estimators, this functional consists of a single summation, whereas for the FRE estimator a more complicated double summation needs to be evaluated. Therefore, for a proper comparison it is better to consider the *complexity* of the different estimators which is defined as the required computation cost for reaching a given accuracy ϵ. For instance, naive evaluation of the double sum in (5) would require a computational cost of order $\mathcal{O}(MN)$ in contrast to $\mathcal{O}(N)$ for the FE and RE estimators. Clearly, such a naive approach would have a serious impact on the complexity of the FRE. Fortunately, smarter procedures for evaluating this double sum exist, which utilize the small support of the kernel K. As a consequence, the main computational cost is due to the simulation of forward and reverse trajectories which is merely order of M+N. For details we refer to Milstein, Schoenmakers, Spokoiny (2004) and also van den Berg, Heemink, Lin, Schoenmakers (2006).

We emphasize that the efficiency of the forward-reverse estimator in comparison with usual ones is essentially based on a specific integral representation for the transition density due to the Chapman-Kolmogorov equation. In this context, the principle of finding efficient estimators has already been used in other respects. Frees (1994) constructs root-N consistent estimators for densities of certain (known) functions of a tuple observations. General results of such type have been obtained by Giné and Mason (2006a,b). Delaigle, Hall and MÍuller (2006) provide another example where root-N consistent estimation is possible in the context of nonparametric regression models, whereas Saavedra and Cao (1999), among others, presented similar results for estimating the stationary density of moving average processes.

Estimator	FE/RE	FRE $d \leq 4$	FRE $d > 4$
δ_N	$N^{-1/(4+d)}$	$N^{-1/d} \log^{1/d} N$	$N^{-2/(4+d)}$
Accuracy	$\mathcal{O}(N^{-2/(4+d)})$	$\mathcal{O}(N^{-1/2})$	$\mathcal{O}(N^{-4/(4+d)})$
Complexity	$\mathcal{O}(\epsilon^{-2-d/2})$	$\mathcal{O}(\lvert\log\varepsilon\rvert\epsilon^{-2})$	$\mathcal{O}(\lvert\log\varepsilon\rvert\epsilon^{-1-d/4})$
$\frac{Compl.\{FE/RE\}}{Compl.\{FRE\}}$		$\lvert\log\varepsilon\rvert^{-1}\epsilon^{-d/2}$	$\lvert\log\varepsilon\rvert^{-1}\epsilon^{-1-d/4}$

Table 1: Summary of accuracy and complexity of the forward (FE), reverse (RE), and forward-reverse (FRE) estimators.

3 Forward probabilistic representations for Markov chains

Consider a discrete-time Markov process $(X_n, \mathcal{F}_n),\ n = 0, 1, 2, ...,$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with phase space $(S, \mathcal{S})$, henceforth called Markov chain. In general we assume that S is locally compact and that $\mathcal{S}$ is the Borel σ-algebra on S. For example, $S = \mathbb{R}^d$ or a discrete subset of $\mathbb{R}^d$. Let $P_n,\ n \geq 0$, denote the one-step transition probabilities defined by

$$P_n(x, B) := \mathbb{P}(X_{n+1} \in B \mid X_n = x),\ n = 0, 1, 2, ...,\quad x \in S,\ B \in \mathcal{S}. \tag{7}$$

In the case of an autonomous Markov chain all the one-step transition probabilities coincide and are equal to $P := P_0 = P_1 = \cdots$.

Let $X_m^{n,x}$, $m \geq n$, be a trajectory of the Markov chain which is at step n in the point x, i.e., $X_n^{n,x} = x$. The multistep transition probabilities $P_{n,m}$ are then defined by

$$P_{n,m}(x, B) := \mathbb{P}(X_m^{n,x} \in B),\quad x \in S,\quad B \in \mathcal{S},\quad m \geq n.$$

Due to these definitions, $P_{n,n}(x, B) = \delta_x(B) = 1_B(x)$ (Dirac measure), $P_n = P_{n,n+1}$, and the Chapman - Kolmogorov equation has the following form:

$$P_{n,m}(x, B) = \int P_{n,k}(x, dy) P_{k,m}(y, B),\qquad x \in S,\ B \in \mathcal{S},\ n \leq k \leq m. \tag{8}$$

Let us fix $N > 0$ and consider for $0 \le n \le N$ the function

$$u_n(x) := \int P_{n,N}(x,dy) f(y) = E\ f(X_N^{n,x}), \tag{9}$$

where f is $\mathcal{S}$-measurable and such that the mathematical expectation in (9) exists; for example, f is bounded. By the Markov property we have for $0 \le n < N$:

$$\begin{aligned} u_n(x) &= E\ f(X_N^{n,x}) = E\ f(X_N^{n+1,X_{n+1}^{n,x}}) \\ &= E\ E^{\mathcal{F}_{n+1}} f(X_N^{n+1,X_{n+1}^{n,x}}) = E\ E^{X_{n+1}^{n,x}} f(X_N^{n+1,X_{n+1}^{n,x}}) \\ &= E\ u_{n+1}(X_{n+1}^{n,x}) = \int u_{n+1}(y) P_n(x,dy). \end{aligned}$$

Thus, $u_n(x)$ satisfies the following discrete integral Cauchy problem

$$u_n(x) = \int u_{n+1}(y) P_n(x,dy),\ n < N, \tag{10}$$

$$u_N(x) = f(x), \tag{11}$$

and (9) is a forward probabilistic representation of its solution. In fact, the probabilistic representation (9) can be used for simulating the solution of (10)-(11) by Monte Carlo.

We now are going to write the discrete Cauchy problem (10)-(11) in another form, thus entailing an alternative probabilistic representation for its solution (9).

Let us consider for $n = 0, 1, ...,$ functions $\rho_n : S \times S \to \mathbb{R}_+$ such that the measures

$$Q_n(x,dy) := \frac{P_n(x,dy)}{\rho_n(x,y)}, \quad x, y \in S \tag{12}$$

are one-step transition functions as well, and functions $h_n : S \times S \to \mathbb{R}$ such that

$$\int h_n(x,y) P_n(x,dy) = 0, \quad x \in S. \tag{13}$$

Note that for arbitrary functions $\widetilde{\rho}_n : S \times S \to \mathbb{R}_+$ the functions $\rho_n(x,y) := \widetilde{\rho}_n(x,y) \int P_n(x,dy')/\widetilde{\rho}_n(x,y')$ satisfy (12), and that for arbitrary functions $\widetilde{h}_n : S \times S \to \mathbb{R}$ the functions $h_n(x,y) := \widetilde{h}(x,y) - \int \widetilde{h}_n(x,y') P_n(x,dy')$ satisfy (13). By (12) and (13), (10)-(11) can be written as

$$u_n(x) = \int (u_{n+1}(y) + h_n(x,y)) \rho_n(x,y) Q_n(x,dy),\ n < N, \tag{14}$$

$$u_N(x) = f(x) \tag{15}$$

The next theorem provides a forward probabilistic representation for the solution of a class of discrete integral Cauchy problems which covers (14)-(15).

Theorem 1 *Let P_n be the one-step transition density of a Markov chain X as in (7) and let the function $f : S \to \mathbb{R}$ be measurable and bounded. Let further $\varphi_n : S \times S \to \mathbb{R}$ and $g_n : S \times S \to \mathbb{R}$ be measurable and bounded functions for $n = 0, 1, 2, ...$ Then, the solution of the problem*

$$\begin{aligned} w_n(x) &= \int (w_{n+1}(z) + g_n(x,z))\varphi_n(x,z)P_n(x,dz), \quad n < N, && (16) \\ w_N(x) &= f(x) && (17) \end{aligned}$$

has the following probabilistic representation:

$$w_n(x) = E\left[f(X_N^{n,x})\mathcal{X}_N^{n,x,1} + \mathbb{X}_N^{n,x,1,0}\right], \tag{18}$$

where $(X, \mathcal{X}, \mathbb{X})$ is an extended Markov chain in which $\mathcal{X}$ and $\mathbb{X}$ are governed by the equations

$$\begin{aligned} \mathcal{X}_{k+1}^{n,x,\gamma} &= \mathcal{X}_k^{n,x,\gamma}\varphi_k(X_k^{n,x}, X_{k+1}^{n,x}), \qquad \mathcal{X}_n^{n,x,\gamma} = \gamma, && (19) \\ \mathbb{X}_{k+1}^{n,x,\gamma,\kappa} &= \mathbb{X}_k^{n,x,\gamma,\kappa} + \varphi_k(X_k^{n,x}, X_{k+1}^{n,x})g_k(X_k^{n,x}, X_{k+1}^{n,x})\mathcal{X}_k^{n,x,\gamma}, \quad \mathbb{X}_n^{n,x,\gamma,\kappa} = \kappa, \end{aligned}$$

where $n \le k < N$.

Proof. Note that $\mathcal{X}_k^{n,x,\gamma} = \gamma\mathcal{X}_k^{n,x,1}$ and $\mathbb{X}_k^{n,x,\gamma,\kappa} = \gamma\mathbb{X}_k^{n,x,1,0} + \kappa$. Thus, for $n < N$, (18) may be written as

$$\begin{aligned} w_n(x) &= E\left[f(X_N^{n+1,X_{n+1}^{n,x}})\mathcal{X}_N^{n+1,X_{n+1}^{n,x},\mathcal{X}_{n+1}^{n,x,1}} + \mathbb{X}_N^{n+1,X_{n+1}^{n,x},\mathcal{X}_{n+1}^{n,x,1},\mathbb{X}_{n+1}^{n,x,1,0}}\right] \\ &= E\ \mathcal{X}_{n+1}^{n,x,1} E^{(X_{n+1}^{n,x},\mathcal{X}_{n+1}^{n,x,1},\mathbb{X}_{n+1}^{n,x,1,0})}\left[f(X_N^{n+1,X_{n+1}^{n,x}})\mathcal{X}_N^{n+1,X_{n+1}^{n,x},1} + \mathbb{X}_N^{n+1,X_{n+1}^{n,x},1,0}\right] \\ &\qquad + E\ \mathbb{X}_{n+1}^{n,x,1,0} \\ &= E\ \left[\mathcal{X}_{n+1}^{n,x,1}w_{n+1}(X_{n+1}^{n,x}) + \mathbb{X}_{n+1}^{n,x,1,0}\right] \\ &= E\ \left[\varphi_n(x, X_{n+1}^{n,x})w_{n+1}(X_{n+1}^{n,x}) + \varphi_n(x, X_{n+1}^{n,x})g_n(x, X_{n+1}^{n,x})\right] \\ &= \int (w_{n+1}(z) + g_n(x,z))\varphi_n(x,z)P_n(x,dz), \end{aligned}$$

and (17) is trivially fulfilled for $n = N$. ■

4 Reverse probabilistic representations

For definiteness we take $S = \mathbb{R}^d$ in this section, hence $\mathcal{S} = \mathcal{B}(\mathbb{R}^d)$, and assume that the transition probabilities $P_{n,m}(x, dy)$ have densities $p_{n,m}(x, y)$ with respect to the Lebesgue measure on $(S, \mathcal{S})$. Then the representation (9) can be written in the form

$$I(f) := E\ f(X_N^{n,x}) = \int p_{n,N}(x,y)f(y)dy. \tag{20}$$

Let the initial value ξ of the chain X at moment n be random with density $g(x)$. Consider the functional

$$I(g,f) = \int\int g(x)p_{n,N}(x,y)f(y)dxdy = Ef(X_N^{n,\xi}). \tag{21}$$

Formally, by taking for g a δ-function we obtain (20) again, and by taking f to be a δ-function we obtain the integral

$$J(g) := \int g(x)p_{n,N}(x,y)dx, \tag{22}$$

We now propose suitable (reverse) probabilistic representations for $J(g)$. From the Chapman - Kolmogorov equation (8) we obtain straightforwardly the Chapman - Kolmogorov equation for densities,

$$p_{n,m}(x,y) = \int p_{n,k}(x,z)p_{k,m}(z,y)dz, \quad x,y \in S, \ n \leq k \leq m, \tag{23}$$

where a "density" $p_{n,n}$ is to be interpreted as a Dirac distribution (δ-function). Let us fix n and N, $n < N$, and introduce the functions

$$v_k(y) := \int g(x)p_{n,k}(x,y)dx, \ n \leq k \leq N, \tag{24}$$

where g is an arbitrary integrable function on S, not necessarily a density. From (23) we get

$$\begin{aligned} v_k(y) &= \int v_{k-1}(z)p_{k-1}(z,y)dz, \ n < k \leq N, \\ v_n(y) &= g(y), \end{aligned} \tag{25}$$

where $p_{k-1} := p_{k-1,k}$ denote the one-step densities. We are going to construct a class of reverse Markov chains which give a probabilistic representation for the solution of (25); hence $J(g)$. For this we introduce for $n < k \leq N$ a reversed time variable $m = N + n - k$ and consider functions $\psi_m : S \times S \to \mathbb{R}_+$ such that for each m and y the function

$$q_m(y,\cdot) := \frac{p_{N+n-m-1}(\cdot,y)}{\psi_m(y,\cdot)} \tag{26}$$

is a density on S. For example, one could take ψ_m independent of the second argument, and then obviously

$$\psi_m(y) = \int p_{N+n-m-1}(z,y)dz. \tag{27}$$

We next introduce $\widetilde{v}_m(y) := v_{N+n-m}(y)$, and transform the problem (25) into

$$\begin{aligned} \widetilde{v}_m(y) &= \int \widetilde{v}_{m+1}(z)\psi_m(y,z)q_m(y,z)dz, \quad n \leq m < N, \\ \widetilde{v}_N(y) &= g(y). \end{aligned} \tag{28}$$

Via Theorem 1 we thus obtain a probabilistic representation of the form (18) for the solution of problem (28), hence (25) and $J(g)$. Indeed, by taking in Theorem 1 instead of X a Markov chain Y, where Y is governed by the one-step transition probabilities $Q_m(y, dz) := q_m(y, z)dz$ (hence Q instead of P), constructing $\mathcal{Y}$ according to

$$\mathcal{Y}_{k+1}^{m,y,1} = \mathcal{Y}_k^{m,y,1}\psi_k(Y_k^{m,y}, Y_{k+1}^{m,y}), \qquad \mathcal{Y}_m^{m,y,1} = 1, \qquad m \leq k < N, \tag{29}$$

and taking $g_n \equiv 0$, it follows by Theorem 1 that

$$\widetilde{v}_m(y) = v_{N+n-m}(y) = E\left[g(Y_N^{m,y})\mathcal{Y}_N^{m,y,1}\right], \qquad n \leq m \leq N, \tag{30}$$

and in particular

$$J(g) = v_N(y) = \int g(x)p_{n,N}(x,y)dx = E\left[g(Y_N^{n,y})\mathcal{Y}_N^{n,y,1}\right]. \tag{31}$$

The representation (31), with $\mathcal{Y}$ given by (29), is a reverse probabilistic representation due to the time reversed chain Y. Obviously, in general different choices for the functions ψ_m give rise to different reverse representations for $J(g)$.

5 Variance reduction

In this section we discuss how to obtain variance reduction in the probabilistic representations (9) and (31). To this aim we consider the variance of the random variable

$$\varsigma := f(X_N^{n,x})\mathcal{X}_N^{n,x,1} + \mathbb{X}_N^{n,x,1,0} \tag{32}$$

in Theorem 1 and prove the next theorem.

Theorem 2 *Let $w_n, P_n,, f, \varphi_n, g_n$ as in Theorem 1. Then it holds*

$$w_k(X_k^{n,x})\mathcal{X}_k^{n,x,1} + \mathbb{X}_k^{n,x,1,0} = E^{(X_k^{n,x}, \mathcal{X}_k^{n,x,1}, \mathbb{X}_k^{n,x,1,0})}\left[w_{k+1}(X_{k+1}^{n,x})\mathcal{X}_{k+1}^{n,x,1} + \mathbb{X}_{k+1}^{n,x,1,0}\right], \tag{33}$$

and

$$\begin{aligned} & Var^{(X_k^{n,x}, \mathcal{X}_k^{n,x,1}, \mathbb{X}_k^{n,x,1,0})}\left[w_{k+1}(X_{k+1}^{n,x})\mathcal{X}_{k+1}^{n,x,1} + \mathbb{X}_{k+1}^{n,x,1,0}\right] \\ = \; & E^{(X_k^{n,x}, \mathcal{X}_k^{n,x,1}, \mathbb{X}_k^{n,x,1,0})} \; (\mathcal{X}_k^{n,x,1})^2(\varphi_k(X_k^{n,x}, X_{k+1}^{n,x})w_{k+1}(X_{k+1}^{n,x}) \\ & + \varphi_k(X_k^{n,x}, X_{k+1}^{n,x})g_k(X_k^{n,x}, X_{k+1}^{n,x}) - w_k(X_k^{n,x}))^2. \end{aligned} \tag{34}$$

As a consequence, if φ_n, g_n, and w_k in (16)-(17) are such that

$$\varphi_k(x, y)(w_{k+1}(y) + g_k(x, y)) = w_k(x), \qquad x, y \in S, \quad n \leq k < N, \tag{35}$$

then

$$w_k(X_k^{n,x})\mathcal{X}_k^{n,x,1} + \mathbb{X}_k^{n,x,1,0} = f(X_N^{n,x})\mathcal{X}_N^{n,x,1} + \mathbb{X}_N^{n,x,1,0}, \qquad n \leq k \leq N, \quad a.s.$$

hence, the random variable (32) is deterministic.

Proof. For $k<N$ we may write using the abbreviation $E^k := E^{(X_k^{n,x},\mathcal{X}_k^{n,x,1},\mathbb{X}_k^{n,x,1,0})}$:

$$\begin{gathered}
w_k(X_k^{n,x})\mathcal{X}_k^{n,x,1}+\mathbb{X}_k^{n,x,1,0}\\
=\mathcal{X}_k^{n,x,1}E^k\left[f(X_N^{k,X_k^{n,x}})\mathcal{X}_N^{k,X_k^{n,x},1}+\mathbb{X}_N^{k,X_k^{n,x},1,0}\right]+\mathbb{X}_k^{n,x,1,0}\\
=E^k\left[f(X_N^{k,X_k^{n,x}})\mathcal{X}_N^{k,X_k^{n,x},\mathcal{X}_k^{n,x,1}}+\mathbb{X}_N^{k,X_k^{n,x},\mathcal{X}_k^{n,x,1},\mathbb{X}_k^{n,x,1,0}}\right]\\
=E^k\left[f(X_N^{k+1,X_{k+1}^{n,x}})\mathcal{X}_N^{k+1,X_{k+1}^{n,x},\mathcal{X}_{k+1}^{n,x,1}}+\mathbb{X}_N^{k+1,X_{k+1}^{n,x},\mathcal{X}_{k+1}^{n,x,1},\mathbb{X}_{k+1}^{n,x,1,0}}\right]\\
=E^k\left[\mathcal{X}_{k+1}^{n,x,1}E^{k+1}\left[f(X_N^{k+1,X_{k+1}^{n,x}})\mathcal{X}_N^{k+1,X_{k+1}^{n,x},1}+\mathbb{X}_N^{k+1,X_{k+1}^{n,x},1,0}\right]+\mathbb{X}_{k+1}^{n,x,1,0}\right]\\
=E^k\left[w_{k+1}(X_{k+1}^{n,x})\mathcal{X}_{k+1}^{n,x,1}+\mathbb{X}_{k+1}^{n,x,1,0}\right],
\end{gathered}$$

thus proving (33). Next, by (33) we have

$$\begin{aligned}
&Var^{(X_k^{n,x},\mathcal{X}_k^{n,x,1},\mathbb{X}_k^{n,x,1,0})}\left[w_{k+1}(X_{k+1}^{n,x})\mathcal{X}_{k+1}^{n,x,1}+\mathbb{X}_{k+1}^{n,x,1,0}\right]\\
=\;&E^k\left[w_{k+1}(X_{k+1}^{n,x})\mathcal{X}_{k+1}^{n,x,1}+\mathbb{X}_{k+1}^{n,x,1,0}-w_k(X_k^{n,x})\mathcal{X}_k^{n,x,1}-\mathbb{X}_k^{n,x,1,0}\right]^2\\
=\;&E^k\left(\mathcal{X}_k^{n,x,1}\right)^2(\varphi_k(X_k^{n,x},X_{k+1}^{n,x})w_{k+1}(X_{k+1}^{n,x})\\
&+\varphi_k(X_k^{n,x},X_{k+1}^{n,x})g_k(X_k^{n,x},X_{k+1}^{n,x})-w_k(X_k^{n,x}))^2,
\end{aligned}$$

hence (34). ■

Let us now go back to the with (9) equivalent Cauchy problem (14)-(15). The solution of this problem has a probabilistic representation according to Theorem 1. Spelling it out, we have

$$u_n(x)=E\left[f(X_N^{n,x})\mathcal{X}_N^{n,x,1}+\mathbb{X}_N^{n,x,1,0}\right], \tag{36}$$

where $\mathcal{X}$ and $\mathbb{X}$ are governed by the equations

$$\begin{aligned}
\mathcal{X}_{k+1}^{n,x,1}&=\mathcal{X}_k^{n,x,1}\rho_k(X_k^{n,x},X_{k+1}^{n,x}), && \mathcal{X}_n^{n,x,1}=1,\\
\mathbb{X}_{k+1}^{n,x,1,0}&=\mathbb{X}_k^{n,x,1,0}+\rho_k(X_k^{n,x},X_{k+1}^{n,x})h_k(X_k^{n,x},X_{k+1}^{n,x})\mathcal{X}_k^{n,x,1}, && \mathbb{X}_n^{n,x,1,0}=0,
\end{aligned}$$

$n\le k<N$. By Theorem 2 the variance of this probabilistic representation vanishes if

$$\rho_k(x,y)(u_{k+1}(y)+h_k(x,y))=u_k(x), \qquad n\le k<N. \tag{37}$$

In principle, (37) may be satisfied while (12) and (13) hold. Indeed, if the functions ρ_k satisfy (12), then the h_k obtained by solving (37) satisfy (13). Vice versa, if the functions h_k satisfy (13) then the ρ_k obtained by solving (37) satisfy (12). In general the exact solution u_k is not known of course, so it will be not possible to choose ρ_k and h_k such that ς in (32) is true deterministic. However, if we have a good approximation $\widehat{u}_k$ of u_k, $n\le k\le N$, at hand, there are possibilities for variance reduction. To formalize the idea of an approximation, we assume that $\widehat{u}_k$ is a known solution of the problem

$$\begin{aligned} \widehat{u}_k(x) &= \int \widehat{u}_{k+1}(y)P_k(x,dy) + \widehat{g}_k(x), \qquad n \le k < N, \\ \widehat{u}_N(x) &= \widehat{f}(x), \end{aligned} \tag{38}$$

where $\widehat{f} - f$ and the $\widehat{g}_k$ are close to zero in some sense.

Importance sampling. Let us assume that both f and its approximation $\widehat{f}$ are positive, and that the approximate solution $\widehat{u}_k$ is for all k positive as well. By then taking $h_k \equiv 0$ for all k, and

$$\rho_k(x,y) = \frac{\widehat{u}_k(x) - \widehat{g}_k(x)}{\widehat{u}_{k+1}(y)}, \tag{39}$$

(12) holds and we may expect that the variance of (32) will be close to zero. If f does not satisfy $f > 0$ but is bounded from below we may shift the problem by choosing a constant C such that $f + C > 0$, and then consider (9) with f replaced by $f + C$.

Control variates. By taking $\rho_k \equiv 1$ for all k, and

$$h_k(x,y) = \widehat{u}_k(x) - \widehat{g}_k(x) - \widehat{u}_{k+1}(y),$$

(13) holds and again we may expect that the variance of (32) will be close to zero.

Combined variance reduction. Importance sampling and control variates can be combined in the following way. Assume that functions $\rho_k > 0$ may be identified such that (12) holds. Then by taking

$$h_k(x,y) = \frac{\widehat{u}_k(x) - \widehat{g}_k(x)}{\rho_k(x,y)} - \widehat{u}_{k+1}(y),$$

(13) holds and we may expect that the Monte Carlo estimator for $u_n(x)$ corresponding to (36) has low variance.

For the reverse probabilistic representation (29)-(31) of the solution of (28) analogue variance reduction methods apply. To be more specific, let $\widetilde{\rho}_m$ and $\widetilde{h}_m$ be such that

$$\widetilde{Q}_m(y,dz) := Q_m(y,dz)/\widetilde{\rho}_m(y,z) = q_m(y,z)dz/\widetilde{\rho}_m(y,z) \tag{40}$$

are one step transition probabilities (see (26)), and

$$\int \widetilde{h}_m(y,z)\psi_m(y,z)Q_m(y,dz) = 0 \tag{41}$$

for all $y \in S$. Then with $\widetilde{\psi}_m := \psi_m\widetilde{\rho}_m$ (28) may be equivalently written as the integral Cauchy problem

$$\begin{aligned} \widetilde{v}_m(y) &= \int (\widetilde{v}_{m+1}(z) + \widetilde{h}_m(y,z))\widetilde{\psi}_m(y,z)\widetilde{Q}_m(y,dz), \quad n \le m < N, \\ \widetilde{v}_N(y) &= g(y) \end{aligned}$$

for which Theorem 1 gives a probabilistic representation of its solution. According to Theorem 2 the variance of the corresponding random variable (32) in this probabilistic representation vanishes when

$$\widetilde{\psi}_m(y,z)(\widetilde{v}_{m+1}(z)+\widetilde{h}_m(y,z))=\widetilde{v}_m(y), \qquad y,z\in S, \quad n\le m<N.$$

Based on an approximate solution of Cauchy problem (28) of the form,

$$\begin{aligned}\widehat{v}_m(y) &= \int \widehat{v}_{m+1}(z)\psi_m(y,z)Q_m(y,dz)+\widehat{a}_m(y), \qquad n\le m<N, \\ \widehat{v}_N(y) &= \widehat{g}(y),\end{aligned} \tag{42}$$

we may apply importance sampling, control variates, or a combination of both as for the forward representation above. For example, if $\widetilde{\rho}_m$ is such that (40) holds, then

$$\widetilde{h}_m(y,z)=\frac{\widehat{v}_m(y)-\widehat{a}_m(y)}{\widetilde{\psi}_m(y,z)}-\widehat{v}_{m+1}(z)$$

satisfies (41) and may lead to a variance reduced representation.

Concluding we may say that Theorem 2 provides variance reduction methods as combinations of importance sampling and control variates and thus can be regarded as a discrete time version of Th. 4.2 in Milstein and Schoenmakers (2002) and Th. 2.1 in Milstein, Schoenmakers, Spokoiny (2004).

6 Reversing autonomous Markov chains

In this section we study Markov chains in the state space $S=\mathbb{R}^d$ with autonomous one step transition density $p(x,y)$ (p does not depend on the step number due to autonomy). If the density $p(x,y)$ and the integral function

$$\psi(y):=\int p(x,y)dx, \qquad y\in\mathbb{R}^d,$$

are known, we can define a Markov chain Y with one step transition density

$$q(y,z)=\frac{p(z,y)}{\psi(y)} \tag{43}$$

and then give for the solution of (24) the following reverse probabilistic representation (see (29)-(31)):

$$\begin{aligned}v_k(y) &= E\left[g(Y_N^{N+n-k,y})\mathcal{Y}_N^{N+n-k,y,1}\right], \\ Y_{N+n-k}^{N+n-k,y} &= y, \ n\le k\le N, \\ \mathcal{Y}_{r+1}^{N+n-k,y,1} &= \mathcal{Y}_r^{N+n-k,y,1}\psi(Y_{r+1}^{N+n-k,y}), \quad N+n-k\le r<N, \\ \mathcal{Y}_{N+n-k}^{N+n-k,y,1} &= 1.\end{aligned} \tag{44}$$

A large class of autonomous Markov chains can be written in the form

$$X_{n+1} = A(X_n, \xi_{n+1}), \;\; n = 0, 1, 2, ..., \tag{45}$$

where ξ_n, $n = 1, 2, ...$, are i.i.d. random variables with density ϕ in $\mathbb{R}^d$. Let us assume that the map $A : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is continuously differentiable and such that there exists a continuously differentiable inverse α with

$$A(x, \alpha(x, z)) = z. \tag{46}$$

For any bounded measurable f the one-step transition density $p(x, y)$ satisfies

$$\begin{aligned} \int f(y)p(x,y)dy &= E\; f(X_{n+1}^{n,x}) = Ef(A(x, \xi_{n+1})) \qquad (47) \\ &= \int f(A(x,\xi))\phi(\xi)d\xi = \int f(y)\phi(\alpha(x,y)) \left|\frac{\partial \alpha(x,y)}{\partial y}\right| dy. \end{aligned}$$

Hence, from (47) it follows that

$$p(x, y) = \phi(\alpha(x, y)) \left|\det \frac{\partial \alpha(x, y)}{\partial y}\right|. \tag{48}$$

According to Section 2, a reverse chain is identified by choosing a function $\psi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$, such that

$$q(y, z) := \frac{p(z, y)}{\psi(y, z)} \tag{49}$$

is a density in z for fixed y. Of course there are infinitely many possible possibilities. The choice (43), for instance, gives the one-step density

$$q(y,z) = \frac{\phi(\alpha(z,y)) \left|\det \frac{\partial \alpha(z,y)}{\partial y}\right|}{\int \phi(\alpha(x,y)) \left|\det \frac{\partial \alpha(x,y)}{\partial y}\right| dx}.$$

When there exists in addition a continuously differentiable inverse β such that

$$A(\beta(z, y), y) = z, \tag{50}$$

we may consider the "physically reversed" chain

$$Y_{n+1} = \beta(Y_n, \widetilde{\xi}_{n+1}), \;\; n = 0, 1, 2, ..., \tag{51}$$

where the sequence $\widetilde{\xi}_n$ is an i.i.d copy of the sequence ξ_n, hence

$$A(Y_{n+1}, \widetilde{\xi}_{n+1}) = A(\beta(Y_n, \widetilde{\xi}_{n+1}), \widetilde{\xi}_{n+1}) = Y_n.$$

The chain (51) has a one-step density of the form (49) for a particular choice of ψ. Indeed, let $q(y, z)$ be the one-step transition density for Y_n. Then, by noting

$\beta(y,\alpha(z,y)) = z$ and writing $\eta(y,z) := \alpha(z,y)$, we have similar to (48),

$$q(y,z) = \phi(\eta(y,z)) \left|\det \frac{\partial \eta(y,z)}{\partial z}\right| = \phi(\alpha(z,y)) \left|\det \frac{\partial \alpha(z,y)}{\partial z}\right| \tag{52}$$
$$= p(z,y) \frac{|\det \partial\alpha(z,y)/\partial z|}{|\det \partial\alpha(z,y)/\partial y|} = p(z,y) \left|\det \frac{\partial A}{\partial x}(z,\alpha(z,y))\right|.$$

Thus, by taking for all k,

$$\psi_k(y,z) := \psi(y,z) := \left|\det \frac{\partial A}{\partial x}(z,\alpha(z,y))\right|^{-1}$$

in the construction of Section 2 and applying Theorem 1, we obtain as in (30) a probabilistic representation for the solution of

$$u_m(y) = v_{N+n-m}(y) = \int g(x) p_{n,N+n-m}(x,y)dx, \qquad n \le m \le N,$$

(hence the integral (24)) of the form

$$\begin{aligned} v_k(y) &= E\left[g(Y_N^{N+n-k,y})\mathcal{Y}_N^{N+n-k,y,1}\right], \qquad n \le k \le N, \\ Y_{r+1}^{N+n-k,y} &= \beta(Y_r^{N+n-k,y},\widetilde{\xi}_{r+1}),\ N+n-k \le r < N, \\ \mathcal{Y}_{r+1}^{N+n-k,y,1} &= \mathcal{Y}_r^{N+n-k,y,1} \left|\det \frac{\partial A}{\partial x}(Y_{r+1}^{N+n-k,y},\alpha(Y_{r+1}^{N+n-k,y},Y_r^{N+n-k,y}))\right|^{-1}. \end{aligned} \tag{53}$$

Example 3 Let A in (45) be of the form,

$$A(x,\xi) = B(x) + C(x)\xi, \quad B:\mathbb{R}^d \to \mathbb{R}^d, \quad C:\mathbb{R}^d \to \mathbb{R}^{d\times d}, \quad \xi \in \mathbb{R}^d. \tag{54}$$

In fact, such a form A arises when discretizing a diffusion SDE and then exchanging Wiener increments with some i.i.d. system of random variables $(\xi_n)_{n=1,2,\ldots}$, which reflect certain features (for example heavy tails) of a problem under consideration. As a special case we consider the Markov chain

$$X_{n+1} = A(X_n,\xi_{n+1}) = BX_n + C\xi_{n+1}, \quad B,C \in \mathbb{R}^{d\times d}, \tag{55}$$

which may be regarded as a discrete Ornstein-Uhlenbeck process under an i.i.d. noise sequence ξ_n, $n = 1,2,\ldots$, of not necessarily Gaussian random variables. Let us suppose that B and C are invertible. Then, α and β in (46) and (50) exist with

$$\alpha(x,z) = C^{-1}(z - Bx), \qquad \beta(z,y) = B^{-1}(z - Cy),$$

and as a reverse chain we may take

$$\begin{aligned}
Y_{k+1}^{m,y} &= B^{-1}(Y_k^{m,y} - C\widetilde{\xi}_{k+1}), \qquad Y_m^{m,y} = y, \\
\mathcal{Y}_{k+1}^{m,y,1} &= \frac{\mathcal{Y}_k^{m,y,1}}{|\det B|}, \qquad m \le k < N, \\
\mathcal{Y}_m^{m,y,1} &= 1, \qquad n \le m \le N.
\end{aligned}$$

In the case where the ξ_n are Gaussian the chains X and Y are Gaussian as well and all characteristics of X and Y can be computed analytically. However, this analytical tractability is generally lost when X is governed by (55) with non-Gaussian ξ_n, for example i.i.d. copies of some heavy tailed distribution.

Example 4 Consider a discrete Black-Scholes model

$$\begin{aligned}
X_{n+1}^i &= A_i(X_n, \xi) := X_n^i \exp\left[\mu_i + (\sigma_i)^\top \xi_{n+1}\right], \\
\mu_i &\in \mathbb{R}, \quad \sigma_i \in \mathbb{R}^d, \quad X_0^i > 0, \quad i = 1, ..., d,
\end{aligned}$$

where the σ_i and ξ_n are to be interpreted as column vectors in $\mathbb{R}^d$, and the $d \times d$ matrix $\sigma := [\sigma_1, \ldots, \sigma_d]^\top$ is assumed to be invertible with inverse σ^{-1}. Then, α and β in (46) and (50) exist, where

$$\begin{aligned}
\alpha_i(x, z) &= \sum_{k=1}^d (\sigma^{-1})_{ik} \left(-\mu_k - \ln x_k + \ln z_k\right), \\
\beta_i(z, y) &= z_i \exp\left[-\mu_i - (\sigma_i)^\top y\right].
\end{aligned}$$

Note that $\partial A_i(x, y)/\partial x_k = \delta_{ik} \exp\left[\mu_i + (\sigma_i)^\top y\right]$, and so after a little algebra we obtain a reverse chain given by

$$\begin{aligned}
Y_{k+1}^{i\,m,y} &= \beta_i(Y_k, \widetilde{\xi}_{k+1}) = Y_k^{i\,m,y} \exp\left[-\mu_i - (\sigma_i)^\top \widetilde{\xi}_{k+1}\right], \\
Y_m^{i\,m,y} &= y_i, \quad i = 1, ..., d, \\
\mathcal{Y}_{k+1}^{m,y,1} &= \mathcal{Y}_k^{m,y,1} \exp\left[-\mu^\top \mathbf{1} - \sum_{i,j} \sigma_{ij} \alpha_j(Y_{k+1}^{m,y}, Y_k^{m,y})\right] = \mathcal{Y}_k^{m,y,1} \prod_{i=1}^d \frac{Y_{k+1}^{i\,m,y}}{Y_k^{i\,m,y}} \\
&= \mathcal{Y}_k^{m,y,1} \exp\left[-\mu^\top \mathbf{1} - \mathbf{1}^T \sigma \widetilde{\xi}_{k+1}\right], \qquad m \le k < N, \\
\mathcal{Y}_m^{m,y,1} &= 1, \qquad n \le m \le N,
\end{aligned}$$

according to (53). For Gaussian noise, both X and Y are log-Gaussian Markov chains, hence analytically tractable like in the previous example. But also here the analytical tractability is lost when non-Gaussian noise is considered.

Example 5 Let us consider the following stochastic volatility model:

$$\begin{aligned}
X_{n+1} &= X_n + f(V_n)\xi_{n+1}, \\
V_{n+1} &= c + g(X_n) + \eta_{n+1},
\end{aligned}$$

where $c > 0$, $f, g : \mathbb{R} \to \mathbb{R}$ are smooth and invertible functions with continuous non-zero derivatives, and $(\xi_n, \eta_n)_{n=1,2,\ldots}$ are i.i.d. random variables. Hence,

$$A(x, v, \xi, \eta) = [x + f(v)\xi, c + g(x) + \eta]^\top$$

and according to (46), (50), we solve α_1, α_2 from

$$\begin{aligned} x + f(v)\alpha_1 &= r \\ c + g(x) + \alpha_2 &= s, \end{aligned}$$

yielding

$$\begin{aligned} \alpha_1(x, v, r, s) &= \frac{r - x}{f(v)}, \qquad f(v) \neq 0, \\ \alpha_2(x, v, r, s) &= s - g(x) - c, \end{aligned}$$

and β_1, β_2 from

$$\begin{aligned} \beta_1 + f(\beta_2)\xi &= y, \\ c + g(\beta_1) + \eta &= w, \end{aligned}$$

yielding

$$\begin{aligned} \beta_1(y, w, \xi, \eta) &= g^{inv}(w - \eta - c) \\ \beta_2(y, w, \xi, \eta) &= f^{inv}(\xi^{-1}y - \xi^{-1}g^{inv}(w - \eta - c)), \qquad \xi \neq 0. \end{aligned}$$

Thus, as a reverse chain we may take

$$\begin{aligned} Y_{k+1}^{m,y,w} &= g^{inv}(W_k^{m,y,w} - \eta_{k+1} - c), \qquad Y_m^{m,y,w} = y \\ W_{k+1}^{m,y,w} &= f^{inv}(\xi_{k+1}^{-1}Y_k^{m,y,w} - \xi_{k+1}^{-1}Y_{k+1}^{m,y,w}), \qquad W_m^{m,y,w} = w \\ \mathcal{Y}_{k+1}^{m,y,w,1} &= \frac{\mathcal{Y}_k^{m,y,w,1}}{\left|g'(Y_{k+1}^{m,y,w})f'(W_{k+1}^{m,y,w})\right|}, \qquad \mathcal{Y}_m^{m,y,w,1} = 1 \qquad n \leq m \leq N. \end{aligned}$$

Note that the inverse $(r, s) \to \alpha(x, v, r, s)$ exist only for $f(v) \neq 0$. This corresponds to the fact that the one step transition density of X does not exist when $f(V_n) = 0$. On the other hand the inverse $(y, w) \to \beta(y, w, \xi, \eta)$ exists only for $\xi \neq 0$. This means that the construction of Y_{k+1} breaks down for a draw $\xi_{k+1} = 0$. However, since the random variables (ξ_n, η_n) are assumed to have a density, and $f' \neq 0$, both $\xi_{k+1} = 0$ and $f(V_n) = 0$ are events of probability zero.

7 Reversing the jump chain of a jump-diffusion

We consider a pure jump process[1] J_t, $t \geq 0$, in $\mathbb{R}$, with jump times $0 < \tau_1 < \tau_2 < \cdots$, where $\tau_{k+1} - \tau_k$ are i.i.d. according to an exponential distribution with

[1] From now on we rather denote time parameters by subscripts.

parameter λ, $\tau_0 := 0$, $J_0 := 0$, and where the jumps are i.i.d. according to a density ν on $\mathbb{R}$. Hence

$$J_t = \sum_{k:\ \tau_{k+1} \leq t} \Delta J_{\tau_k}$$

with $\Delta J_{\tau_k} := J_{\tau_{k+1}} - J_{\tau_k}$ being i.i.d distributed with density ν. The process J_t is piece-wise constant and is continuous from the right with limits from the left (c.r.l.l.). The solution of the SDE

$$dX_t = \mu(X_{t-}, J_{t-})dt + \sigma(X_{t-}, J_{t-})dW_t + \eta(X_{t-}, J_{t-})dJ_t, \qquad X_0 = x_0,$$

for smooth functions $\mu, \eta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, with a from J independent Wiener process W_t, is generally called a jump diffusion. The process X_t is c.r.l.l. as well and at jump times we have

$$\begin{aligned} X_{\tau_{k+1}} &= X_{\tau_k} + \int_{\tau_k}^{\tau_{k+1}} \mu(X_s, J_{\tau_k})ds + \sigma(X_s, J_{\tau_k})dW_s \\ &\quad + \eta(X_{\tau_{k+1}-}, J_{\tau_k})\Delta J_{\tau_k}, \end{aligned}$$

We now consider the autonomous Markov chain $\mathbf{X}$ in $\mathbb{R}^3$ defined by $\mathbf{X}_k = (\tau_k, X_{\tau_k}, J_{\tau_k})$, $k = 0, 1, ...$, and its associated reverse representations. As a motivation, we imagine that we consider a process of which only its jumps and jump times are of importance. For example, a jump might be connected with some default event and an insurance company is concerned with the occurrence of the10th default, since she then has to pay money due to a certain insurance contract. If the actual default time of the 10th default is irrelevant for the insurance company, one could only consider the autonomous Markov chain (X_{τ_k}, J_{τ_k}), $k = 0, 1, ...$ In many cases the one-step transition density of the latter chain is less tractable, however.

Let $\pi^{a,x}(t, \cdot)$ be the transition density in $\mathbb{R}$ of the process $U_t^{a,x}$ defined as solution of the SDE

$$dU_t := \mu(U_t, a)ds + \sigma(U_t, a)dW_t, \qquad t \geq 0,\ U_0 = x,$$

and let $\alpha(a, b, y)$ be the solution of the equation

$$\alpha + \eta(\alpha, a)(b - a) = y.$$

Then given $\tau_k = \varsigma$, $X_\varsigma = x$, $J_\varsigma = a$, hence $\mathbf{X}_k = (\varsigma, x, a)$, the one-step density of $\mathbf{X}_{k+1} = (\tau_{k+1}, X_{\tau_{k+1}}, J_{\tau_{k+1}})$ in (τ, y, b), $\varsigma < \tau$, is given by

$$p(\varsigma, x, a, \tau, y, b) = \lambda e^{-\lambda(\tau - \varsigma)} \pi^{a,x}(\tau - \varsigma, \alpha(a, b, y))\nu(b - a), \qquad \varsigma < \tau. \tag{56}$$

Clearly, the chain $\mathbf{X}$ is autonomous. By taking the integral

$$\begin{aligned} \psi(\tau, y, b) &= \int_0^\tau \lambda e^{-\lambda(\tau - \varsigma)} d\varsigma \int dx \int \pi^{a,x}(\tau - \varsigma, \alpha(a, b, y))\nu(b - a)da \\ &= \int_0^\tau \lambda e^{-\lambda\varsigma} d\varsigma \int dx \int \pi^{a,x}(\varsigma, \alpha(a, b, y))\nu(b - a)da, \end{aligned} \tag{57}$$

we obtain from (43) the one-step transition density

$$
\begin{aligned}
q(\tau, y, b, \vartheta, z, c) &:= \frac{p(\vartheta, z, c, \tau, y, b)}{\psi(\tau, y, b)} \\
&= \frac{\lambda e^{-\lambda(\tau-\vartheta)} \pi^{c,z}(\tau-\vartheta, \alpha(c, b, y)) \nu(b-c)}{\psi(\tau, y, b)}, \quad \vartheta < \tau, \quad (58)
\end{aligned}
$$

and the corresponding reverse Markov chain $(\mathbf{Y}, \mathcal{Y})$ in $\mathbb{R}^3 \times \mathbb{R}$, where $\mathbf{Y}_k := (\Theta_k, Z_k, C_k)$ is governed by the one-step density (58), with

$$
\Theta_m^{m,\ \tau,y,b} = \tau,\ Z_m^{m,\ \tau,y,b} = y,\ C_m^{m,\ \tau,y,b} = b, \quad (59)
$$

and $\mathcal{Y}$ satisfies

$$
\begin{aligned}
\mathcal{Y}_{k+1}^{m,\ \tau,y,b,1} &= \mathcal{Y}_k^{m,\ \tau,y,b,1} \psi(\Theta_k^{m,\ \tau,y,b}, Z_k^{m,\ \tau,y,b}, C_k^{m,\ \tau,y,b}), \\
\mathcal{Y}_m^{m,\ \tau,y,b,1} &= 1, \quad m \le k < N, \quad (60)
\end{aligned}
$$

(see (29)-(31) and (44)).

As a more particular case we consider the autonomous process

$$
dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dJ_t, \quad (61)
$$

where the diffusion component of X has transition density $\pi^x(t, \cdot)$, which is independent of a. Then, $\mathbf{X}_k := (\tau_k, X_{\tau_k})$, $k = 0, 1, ...$, is an autonomous Markov chain itself, with one step transition density

$$
p(\varsigma, x, \tau, y) = \lambda e^{-\lambda(\tau-\varsigma)} \int \pi^x(\tau-\varsigma, y-u)\nu(u)du, \qquad \varsigma < \tau. \quad (62)
$$

Thus, according to (57), (58), we obtain a reverse chain $\mathbf{Y}$ with one step transition density

$$
q(\tau, y, \vartheta, z) = \frac{p(\vartheta, z, \tau, y)}{\psi(\tau, y)}, \qquad \vartheta < \tau, \quad (63)
$$

where

$$
\psi(\tau, y) = \int_0^\tau d\varsigma \lambda e^{-\lambda\varsigma} \int dx \int \pi^x(\varsigma, y-u)\nu(u)du. \quad (64)
$$

Example 6 Consider the jump-diffusion

$$
dX_t = \kappa(\chi - X_t)dt + \sigma dW_t + dJ_t,
$$

for which the diffusion component

$$
dU_t = \kappa(\chi - U_t)dt + \sigma dW_t
$$

is a mean reverting Gaussian process with transition density

$$
\pi(t, x, u) = \frac{1}{\sqrt{\pi\sigma^2(1-e^{-2\kappa t})/\kappa}} \exp\left[-\frac{(\chi + (x-\chi)e^{-\kappa t} - u)^2}{\sigma^2(1-e^{-2\kappa t})/\kappa}\right],
$$

and so (62) becomes

$$
\begin{aligned}
p(\varsigma,x,\tau,y) &= \frac{\lambda e^{-\lambda(\tau-\varsigma)}}{\sqrt{\pi\sigma^2(1-e^{-2\kappa(\tau-\varsigma)})/\kappa}}\times \\
&\times\int\exp\left[-\frac{(\chi+(x-\chi)e^{-\kappa(\tau-\varsigma)}+u-y)^2}{\sigma^2(1-e^{-2\kappa(\tau-\varsigma)})/\kappa}\right]\nu(u)du,\quad \zeta<\tau.
\end{aligned}
$$

It is easy to see that $\int\pi^x(t,u)dx=e^{\kappa t}$, hence (64) becomes,

$$
\begin{aligned}
\psi(\tau,y) &= \int_0^\tau d\varsigma\lambda e^{-\lambda\varsigma}\int dx\int\pi(\varsigma,x,y-u)\nu(u)du \\
&= \int_0^\tau d\varsigma\lambda e^{-\lambda\varsigma}e^{\kappa\varsigma}\int\nu(u)du=\lambda\frac{e^{(\kappa-\lambda)\tau}-1}{\kappa-\lambda},
\end{aligned}
$$

and the reverse transition density (63) is thus given by

$$
q(\tau,y,\vartheta,z)=\frac{(\kappa-\lambda)\,e^{-\lambda(\tau-\vartheta)}}{\left(e^{(\kappa-\lambda)\tau}-1\right)\sqrt{\pi\sigma^2(1-e^{-2\kappa(\tau-\vartheta)})/\kappa}}\times \tag{65}
$$

$$
\times\int\exp\left[-\frac{(\chi+(z-\chi)e^{-\kappa(\tau-\vartheta)}+u-y)^2}{\sigma^2(1-e^{-2\kappa(\tau-\vartheta)})/\kappa}\right]\nu(u)du,\qquad \vartheta<\tau.
$$

The dynamics (58)-(60) thus collapses to the chain $\mathbf{Y}_k=(\Theta_k,Z_k)$ governed by (65) with

$$
\Theta_m^{m,\ \tau,y}=\tau,\ Z_m^{m,\ \tau,y,b}=y,
$$

and scalar process $\mathcal{Y}$ satisfying

$$
\begin{aligned}
\mathcal{Y}_{k+1}^{m,\ \tau,y,1} &= \mathcal{Y}_k^{m,\ \tau,y,1}\psi(\Theta_k^{m,\ \tau,y},Z_k^{m,\ \tau,y}), \\
\mathcal{Y}_m^{m,\ \tau,y,b,1} &= 1,\qquad m\le k<N.
\end{aligned}
$$

Finally we note that for a variety of jump densities ν, (65) can be expressed in close form (for instance when v is Gaussian).

8 Reversed representations for ODE processes driven by continuous time Markov chains

To complete the picture we describe in this section the reversion of processes obtained as the solution of an ordinary differential equation which is driven by a continuous time Markov chain with finite state space. Suppose we are given a regular Markov chain $(X_t)_{t\ge0}$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ with state space $S=\{x_1,...,x_m\}$. In connection with the chain X we consider a system of ordinary differential equations

$$
\frac{dY}{ds}=a(X,Y), \tag{66}
$$

where $Y = [Y^1, ..., Y^d]^\intercal$, and $a : S \times \mathbb{R}^d \ni (x, y) \to [a^1(x, y), ..., a^d(x, y)]^\intercal \in \mathbb{R}^d$. We assume the functions $a^i_j(y) := a^i(x_j, y)$, $i = 1, ..., d$, $j = 1, ..., m$, have bounded continuous derivatives. Let

$$Q = \begin{bmatrix} -q_1 & q_{12} & \cdots & q_{1m} \\ q_{21} & -q_2 & \cdots & q_{2m} \\ \cdot & \cdot & \cdot & \cdot \\ q_{m1} & q_{m2} & \cdots & -q_m \end{bmatrix},$$

be the infinitesimal generator matrix of the chain X with $q_{ij} \geq 0$, $i \neq j$, and

$$\sum_{j \neq i} q_{ij} = q_i. \tag{67}$$

It is well known that the infinitesimal generator of the Markov process (X, Y) generated by the system (66) is given by

$$Af(x_i, y) = \sum_{k=1}^{d} a^k(x_i, y) \frac{\partial f}{\partial y^k}(x_i, y) - q_i f(x_i, y) + \sum_{j \neq i} q_{ij} f(x_j, y), \tag{68}$$
$$i = 1, ..., m, \ y \in \mathbb{R}^d$$

(see for example Milstein & Repin (1969)). In particular, for some smooth function $f(x, y)$ with bounded derivative with respect to y, the functions $u_i(t, y) := Ef(X^{x_i}(t), Y^{x_i,y}(t))$ satisfy the hyperbolic system

$$\begin{aligned} \frac{\partial u_i}{\partial t}(t, y) &= \sum_{k=1}^{d} a^k(x_i, y) \frac{\partial u_i}{\partial y^k}(x_i, y) - q_i u_i + \sum_{j \neq i} q_{ij} u_i, \\ u_i(0, y) &= f(x_i, y) \quad i = 1, ..., m, \quad y \in \mathbb{R}^d. \end{aligned}$$

By considering the semigroup generated by the extended Markov process $(X, Y, \mathcal{Y})$ on $S \times \mathbb{R}^d \times \mathbb{R}$, where $\mathcal{Y}$ is defined by

$$\frac{d\mathcal{Y}^{x_i,y,\varsigma}}{dt} = b(X, Y)\mathcal{Y}^{x_i,y,\varsigma}, \qquad \mathcal{Y}_0^{x_i,y,\varsigma} = \varsigma,$$

for some bounded real valued function $b(x, y)$ in $S \times \mathbb{R}^d$ with continuous derivatives in y, it follows that the functions

$$v_i(t, y) := E\ f(X_t^{x_i}, Y_t^{x_i,y}) \exp\Big(\int_0^t b(X_s^{x_i}, Y_s^{x_i,y}) ds\Big), \tag{69}$$
$$i = 1, ..., m, \ y \in \mathbb{R}^d, \ t > 0,$$

satisfy the more general hyperbolic system

$$\frac{\partial v_i}{\partial t} = \sum_{k=1}^{d} a^k(x_i, y) \frac{\partial v_i}{\partial y^k} + b(x_i, y) v_i - q_i v_i + \sum_{j \neq i} q_{ij} v_j \tag{70}$$
$$u_i(0, y) = f(x_i, y), \ i = 1, ..., m, \ y \in \mathbb{R}^d.$$

So, (69) is a probabilistic representation for the solution of (70) which may be used for evaluating the $v_i(t,y)$ in (70) by Monte Carlo simulation.

Let an initial distribution of (X,Y) be given by

$$P\{X_0 = x_i,\ Y_0 \in H\} = \int_H \lambda_i(y)dy \tag{71}$$

for $i = 1, ..., m$, with $\lambda_i(y)$ being a density on $\mathbb{R}^d$. Suppose further that the $\lambda_i(y)a^k(x_i,y)$ have continuous partial derivatives $\partial(\lambda_i a_i^k)/\partial y^k$ and that the integrals

$$I_i = \int_{\mathbb{R}^d} \sum_{k=1}^{d} \frac{\partial(\lambda_i a_i^k)}{\partial y^k}(y)dy,\ i = 1, ..., m,$$

exist. Then for each $i = 1, ..., m$, the function $\psi_i(t,y)$ defined by the property

$$\int_H \psi_i(t,y)dy = \mathbb{P}(\{X_t = x_i,\ Y_t \in H\}),$$

satisfies the forward Kolmogorov equation

$$\frac{\partial \psi_i}{\partial t} = -\sum_{k=1}^{d} \frac{\partial(\psi_i a_i^k)}{\partial y^k}(y) - q_i\psi_i + \sum_{j\neq i} q_{ji}\psi_j \tag{72}$$

with initial condition

$$\psi_i(0,y) = \lambda_i(y) = \lambda(x_i,y),\ i = 1, ..., m. \tag{73}$$

In fact $\psi_i(t,\cdot)/\mathbb{P}(\{X(t) = x_i\})$ is the conditional density of Y at time $t > 0$, given $X(t) = x_i$. In order to obtain a probabilistic representation for $\psi_i(t,y)$ we will cast the system (72) into the form of (70):

$$\frac{\partial \psi_i}{\partial t} = -\sum_{k=1}^{d} a^k(x_i,y)\frac{\partial \psi_i}{\partial y^k} + c(x_i,y)\psi_i - q_i^*\psi_i + \sum_{j\neq i} q_{ij}^*\psi_j, \tag{74}$$

where

$$\begin{aligned} q_{ij}^* &= q_{ji},\ i \neq j, \\ q_i^* &= \sum_{j\neq i} q_{ij}^* = \sum_{j\neq i} q_{ji}, \\ c(x_i,y) &= -\sum_{k=1}^{d} \frac{\partial a^k(x_i,y)}{\partial y^k} + q_i^* - q_i. \end{aligned} \tag{75}$$

Thus, for the solution of (72) we obtain a probabilistic representation

$$\psi_i(t,y) = E\lambda(X_t^{*x_i}, Y_t^{*x_i,y})\exp(\int_0^t c(X_s^{*x_i}, Y_s^{*x_i,y})ds), \tag{76}$$

via a reversed process (X^*, Y^*), where X^* is a Markov chain with generator matrix $Q^* := \{q^*_{ij}\}$ with $q^*_{ii} = -q^*_i$, and Y^* is governed by the equation

$$\frac{dY^*}{ds} = -a(X^*, Y^*). \tag{77}$$

Note that the representation (76) for the solution of the problem (72)-(73) holds for any λ with bounded derivatives of first order, hence not only for densities.

Some bibliographical notes. The first probabilistic representation of solutions for hyperbolic equations (for telegraph equation) goes back to M. Kac. Sufficiently general systems of ordinary differential equations driven by a Markov chain were considered in Kac & Krasovsky (1960). The detailed description of the process (X, Y) was done in Milstein & Repin (1969). In a lot of papers such processes were treated in connection with random evolution (see, for instance, Griego & Hersch (1969), Hersch & Papanicolaou (1972) and references therein). In all these papers the process X does not depend on the state of Y. The interaction of general processes X and Y is considered in Milstein (1972). Instead of the system of ordinary differential equations (66) it is possible to examine a system of stochastic differential equations interacting with a Markov chain as well. Both Cauchy problems and boundary value problems for systems of partial differential equations arising in connection with interacting Markov processes are considered in Milstein (1978).

9 Forward-reverse transition density estimation with applications

In this section we describe for a Markov Chain (7) an efficient procedure for estimating the transition density $p_{n,m}(x, y)$ by a forward-reverse probabilistic representation. The procedure below is in fact a discrete-time version of the method developed in Milstein, Schoenmakers, Spokoiny (2004) for continuous-time processes given by an Ito SDE. Thus, let us take x, y and n, m fixed and concentrate on the problem of estimating $p_{n,m}(x, y)$. If $m = n + 1$ this is a one step transition density which is assumed to be given. Therefore we assume $m - n \gg 1$ and, in the spirit of Milstein, Schoenmakers, Spokoiny (2004), for some fixed k^* with $n < k^* < m$ we consider the Chapman - Kolmogorov equation

$$p_{n,m}(x, y) = \int p_{n,k^*}(x, z) p_{k^*,m}(z, y) dz \tag{78}$$

and observe by Section 4 that for $h^x(z) := p_{n,k^*}(x, z)$ (78) has a probabilistic representation

$$p_{n,m}(x, y) = \int h^x(z) p_{k^*,m}(z, y) dz = E\left[h^x(Y_m^{k^*,y}) \mathcal{Y}_m^{k^*,y,1}\right],$$

where $(Y,\mathcal{Y})$ is constructed as in (30). We next consider $\widehat{h}^x$ to be a Parzen-Rozenblatt estimator for the density function $z \to h^x(z)$. Hence,

$$\widehat{h}^x(z) = \widehat{p}_{n,k^*}(x,z) = \frac{1}{L\delta^d}\sum_{l=1}^{L} K(\frac{X^{n,x}_{k^*(l)} - z}{\delta}),$$

where K is some kernel, δ is a bandwidth parameter, and $X^{n,x}_{k^*(l)}$ are independent realisations of $X^{n,x}_{k^*}$ for $l = 1, ..., L$. By replacing $p_{n,k}$ in (78) by its estimator $\widehat{h}$ we obtain

$$\widehat{p}_{n,m}(x,y) := \int \widehat{h}^x(z) p_{k^*,m}(z,y)dz = E\left[\widehat{h}^x(Y^{k^*,y}_m)\mathcal{Y}^{k^*,y,1}_m\right],$$

and completely similar to a proof in Milstein, Schoenmakers, Spokoiny (2004) it follows that the estimator

$$\begin{aligned} \widehat{\widehat{p}}_{n,m}(x,y) \quad &:= \quad \frac{1}{R}\sum_{r=1}^{R} \widehat{h}^x(Y^{k^*,y}_{m(r)})\mathcal{Y}^{k^*,y,1}_{m(r)} \\ &= \quad \frac{1}{LR\delta^d}\sum_{r=1}^{R}\sum_{l=1}^{L} K(\frac{X^{n,x}_{k^*(l)} - Y^{k^*,y}_{m(r)}}{\delta})\mathcal{Y}^{k^*,y,1}_{m(r)}, \end{aligned} \tag{79}$$

with $(Y^{k^*,y}_{m(r)}, \mathcal{Y}^{k^*,y,1}_{m(r)})$ being independent realisations of $(Y^{k^*,y}_m, \mathcal{Y}^{k^*,y,1}_m)$ for $r = 1, ..., R$, is a root-L consistent estimator for the target density $p_{n,m}(x,y)$ when $L = R$. For a detailed study of the properties of the forward-reverse density estimator (79) we refer to Milstein, Schoenmakers, Spokoiny (2004). Below we consider some applications of the forward-reverse estimator (79).

Estimating the probability of visiting a bounded region

Let μ_G be a probability measure concentrated on G. Let $y_1, ..., y_R$ be i.i.d. drawings from μ_G. Then the estimator

$$\widehat{\widehat{p}}_{n,m}(x,\mu_G) := \frac{1}{LR\delta^d}\sum_{l=1}^{L}\sum_{r=1}^{R} K(\frac{X^{n,x}_{k^*(l)} - Y^{k^*,y_r}_{m(r)}}{\delta})\mathcal{Y}^{k^*,y_r,1}_{m(r)} \tag{80}$$

has expectation

$$
\begin{aligned}
E\widehat{\widehat{p}}_{n,m}(x,\mu_G) &:= EE\left[\frac{1}{LR\delta^d}\sum_{l=1}^{L}\sum_{r=1}^{R}K(\frac{X_{k^*(l)}^{n,x}-Y_{m(r)}^{k^*,y_r}}{\delta})\mathcal{Y}_{m(r)}^{k^*,y_r,1}|y_1,...,y_M\right] \\
&= E\frac{1}{R}\sum_{r=1}^{R}E\left[\frac{1}{L\delta^d}\sum_{l=1}^{L}K(\frac{X_{k^*(l)}^{n,x}-Y_{m(r)}^{k^*,y_r}}{\delta})\mathcal{Y}_{m(r)}^{k^*,y_r,1}|y_r\right] \\
&=: E\sum_{r=1}^{R}\frac{1}{R}(p_{n,m}(x,y_r)+\epsilon(\delta,y_r)) \\
&= \int\mu_G(dy)p_{n,m}(x,y)+\int\mu_G(dy)\epsilon(\delta,y) \\
&=: \int\mu_G(dy)p_{n,m}(x,y)+\epsilon(\delta,\mu_G),
\end{aligned}
$$

since the bias of the FRE is independent of L and R. If the kernel K is of sufficient order (for $d \leq 4$ first order is enough) and the region G is bounded, it follows from Section 6 of Milstein, Schoenmakers, Spokoiny (2004) that $\epsilon(\delta,\mu_G) = O(\delta^2)$, and moreover if $(R = L)$,

$$
\mathrm{V}ar\,\widehat{\widehat{p}}_{n,m}(x,\mu_G) \leq \frac{C}{L}+o(\frac{1}{L}), \quad L\to\infty,
$$

where C is a constant.

Suppose G is some bounded (Borel) region in $\mathbb{R}^d$ with small probability to be visited by the chain X in $m-n$ steps when X starts from x at time n. Hence $P_{n,m}(x,G) := \int_G p_{n,m}(x,y)dy$ is small. In such a situation one could estimate $P_{n,m}(x,G)$ in principle with root-N accuracy using a standard Monte Carlo estimator, $\widehat{I}_G$ say, for the probabilistic representation (20), where f is taken to be the indicator of G. However, the *relative* accuracy $\sqrt{\mathrm{V}ar\hat{I}_G}/P_{n,m}(x,G)$ of this estimator is equal to $\sqrt{(P_{n,m}^{-1}(x,G)-1)/N}$. Hence the relative accuracy is root-N, but with a large order coefficient when $P_{n,m}(x,G)$ is small. The FRE estimator (80) mends this problems. Indeed, take μ to be the uniform distribution on G, then (80) yields an estimator for $P_{n,m}(x,G)/\lambda(G)$ (with λ denoting the Lebesgue measure on $\mathbb{R}^d$) with accuracy of $\sqrt{C/N+\epsilon^2(\delta,\mu_G)}$. Hence, $\lambda(G)\widehat{\widehat{p}}_{n,m}(x,\mu_G)$ is an estimator for $P_{n,m}(x,G)$ with relative accuracy $\lambda(G)P_{n,m}^{-1}(x,G)\sqrt{C/N+\epsilon^2(\delta,\mu_G)}$. Since $P_{n,m}(x,G)/\lambda(G)$ can be interpreted as the average density over the region G, the root-N coefficient does not explode when, for instance, the density $p_{n,m}(x,y)$ is continuous and positive for all y.

The problem of estimating the probability of visiting a critical region at a certain time has many applications in the area of environmental modelling, e.g, see Spivakovskaya et al. (2005), van den Berg et al. (2006).

Probability of visiting an unbounded domain

If the domain G is unbounded but such that it does not intersect with a certain sphere, we may map G to the inside of this sphere and then work with the image of the chain X under this map. Spelling it out, let $G \subset \{x \in \mathbb{R}^d : |x - x_0| > r_0\}$ for some $x_0 \in \mathbb{R}^d$ and $r_0 > 0$, and let

$$S : x \to x_0 + \frac{r_0^2}{|x - x_0|^2}(x - x_0), \quad x \in \mathbb{R}^d, \quad x \neq x_0,$$

be the spherical inversion with respect to the sphere of radius r_0 with center x_0. The map S is a bijective transformation on $\mathbb{R}^d \backslash \{x_0\}$ with inverse $S^{-1} = S$, which maps $\{x \in \mathbb{R}^d : |x - x_0| > r_0\}$ onto $\{x \in \mathbb{R}^d : 0 < |x - x_0| < r_0\}$. Let us define $U_m^{n,u} := S(X_m^{n,S^{-1}(u)})$, $u \neq x_0$. Then, clearly,

$$\begin{aligned} P_{n,m}(x, G) &= P(X_m^{n,x} \in G) = P(S(X_m^{n,x}) \in S(G)) \\ &=: P(U_m^{n,S(x)} \in S(G)), \quad x \neq x_0, \end{aligned}$$

with $S(G) := \{S(x) \in \mathbb{R}^d : |x| \in G\}$ being bounded. We so can apply our forward-reverse methodology to the Markov chain U and the region $S(G)$. For this we need to find the one-step transition density of the chain U, denoted by p_n^U. It holds,

$$\begin{aligned} &\int_A p_n^U(u, v)dv := P(U_{n+1}^{n,u} \in A) = P(X_{n+1}^{n,S^{-1}(u)} \in S^{-1}(A)) \\ &= \int_{S^{-1}(A)} p_n(S^{-1}(u), y)dy = \int_A p_n(S^{-1}(u), S^{-1}(v)) \left| \det \frac{\partial S^{-1}}{\partial v} \right| dv, \quad u \neq x_0, \end{aligned}$$

by the standard transformation theorem. We thus yield,

$$p_n^U(u, v) = p_n(S^{-1}(u), S^{-1}(v)) \left| \det \frac{\partial S^{-1}}{\partial v} \right| = p_n(S(u), S(v)) \left| \det \frac{\partial S}{\partial v} \right|, \quad u \neq x_0.$$

Estimating Value at Risk

Let $L(z)$ be the loss (in absolute value) of a portfolio as function of the asset position vector $z = X_m^{n,x}$, and consider $G_a := \{z \in \mathbb{R}^d : L(z) > a\}$ for $a > 0$. Then G_a is typically unbounded but may be contained in the complement of some sphere, so that we may use the above transformation. The value of a such that $P_{n,m}(x, G_a) = \alpha$, where α is a given quantile, e.g. 5%, is called the α% Value at Risk.

Generally, straightforward Monte Carlo evaluation of the Value at Risk of a large portfolio is very time consuming since one needs many sample trajectories to generate a reliable number in a certain critical region. A forward-reverse approach may therefore be considered as an elegant solution.

References

[1] van den Berg, E., Heemink, A.W., Lin, H.X., Schoenmakers, J.G.M., Probability density estimation in stochastic environmental models using reverse representations. *Serra*, **20**, (1-2) (2006), 126–139.

[2] Delaigle, A., Hall, P. and Müller, H.-G., Accelerated convergence for nonparametric regression with coarsened predictors. Working paper (2006), http://www.math.ucsd.edu/~adelaigl/

[3] Devroye, L. and Györfi, L., *Nonparametric Density Estimation: The L_1 View.* Wiley 1985.

[4] Frees, E. W., Estimating densities of functions of observations. *J. Amer. Statist. Assoc.* 89 (1994), 517Ű525.

[5] Giné, E. and Mason, D., On local U-statistic processes and the estimation of densities of functions of several variables. Working paper (2006a), http://www.math.uconn.edu/~gine/, to appear in: *Ann. Statist.*

[6] Gint́e, E. and Mason, D. (2006b). Laws of the iterated logarithm for the local U-statistic process. Working paper (2006b), http://www.math.uconn.edu/~gine/, to appear in: *J. Theoret. Probab.*

[7] Griego, R.J., Hersh, R., Random evolutions, Markov chains and systems of partial differential equations. Proc. Nat. Acad. Sci. USA, **62**, (2) (1969), 305–308.

[8] Hersh, R., Papanicolaou, G., Non-commuting random evolutions and an operator-valued Feynman- Kac formula. Comm.Pure and Appl. Math., **25**, (3) (1972), 337–367.

[9] Kac, I., Ja., Krasovsky, N.N., On stability of systems with random parameters. Appl. Math. Mech., **24**, (5)(1960), 809–823.

[10] Kloeden, P.E., Platen, E., *Numerical solution of stochastic differential equations.* Springer Verlag Berlin, 1992.

[11] Milstein, G.N., Interaction of Markov processes. Theory Prob. Appl., **17**, (1972), 36–45.

[12] Milstein, G.N., Probabilistic solution of linear systems of elliptic and parabolic equations. Theory Prob. Appl., **23** (1978), 851-855.

[13] Milstein, G.N., Repin, Yu.M., The action of a Markov process on a system of differential equations. Differential equations, **5** (1969), 1371-1384.

[14] Milstein, G.N. and Schoenmakers, J.G.M., Monte Carlo construction of hedging strategies against multi-asset European claims. *Stochastics and Stochastics Reports*, **73**, (1-2) (2002), 125–157.

[15] Milstein, G.N., Schoenmakers, J.G.M., Spokoiny, V., Transition density estimation for stochastic differential equations via forward-reverse representations. *Bernoulli*, **10**, no. 2 (2004), 281–312.

[16] Milstein, G.N., Tretyakov, M.V., *Stochastic Numerics for Mathematical Physics*, Springer Verlag Berlin, 2004.

[17] Saavedra, A. and Cao, R., Rate of convergence of a convolutiontype estimator of the marginal density of an MA(1) process. *Stochastic Process. Appl.* **80** (1999), 129Ũ155.

[18] Scott, D.W., *Multivariate Density Estimation.* Wiley, New York, 1992.

[19] Silverman, B.W., *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.

[20] Spivakovskaya, D., Heemink, A.W., Milstein, G.N., Schoenmakers, J.G.M., Simulation of the transport of particles in coastal waters using forward and reverse time diffusion. *Advances in Water Resources*, 28 (2005), 927–938.

[21] Spivakovskaya, D., Heemink, A.W., Schoenmakers, J.G.M., Two-particle models for the estimation of the mean and standard deviation of concentrations in coastal waters. *Stochastic Environmental Research and Risk Assessment (SERRA)* (2006) DOI 10.1007/s00477-006-0059-0, in print.