On generating Monte Carlo samples of continuous diffusion bridges

Ming Lin, Rong Chen, and Per Mykland

Xiamen University, Rutgers University,

Peking University and University of Chicago

March 12, 2010

Abstract

Diffusion processes are widely used in engineering, fiance, physics and other fields. Usually continuous time diffusion processes are only observable at discrete time points. For many applications, it is often useful to impute continuous time bridge samples that follow the diffusion dynamics and connect each pair of the consecutive observations. The Sequential Monte Carlo (SMC) method is a useful tool to generate the intermediate paths of the bridge. Often the paths are generated forward from the starting observation and forced in some ways to connect with the end observation. In this paper we propose a constrained SMC algorithm with an effective resampling scheme that is guided by backward pilots carrying the information of the end observation. This resampling

scheme can be easily combined with any forward SMC sampler. Two synthetic examples are used to demonstrate the effectiveness of the resampling scheme.

1 Introduction

Diffusion processes are widely used in engineering, finance, physics and many other fields. In practice a diffusion process is often only observable at discrete time points. On the other hand, for most nonlinear and non-Gaussian diffusion processes, statistical inferences can be much more easily carried out with continuous paths. Treating the problem as a missing data problem, an effective solution for statistical inferences is to impute the continuous path based on the observations observed at discrete time points. Due to the Markovian nature of diffusion processes, the imputation problem becomes one of generating continuous paths of the underlying diffusion process that connect two fixed end-points (diffusion bridges). In this paper we propose a constrained sequential Monte Carlo (CSMC) algorithm with resampling guided by backward pilots for efficient generation of Monte Carlo samples of diffusion bridges.

Let a d-dimensional time-homogeneous diffusion process V_t be the solution of a stochastic diffusion equation (SDE)

$$dV_t = \boldsymbol{b}(V_t; \theta)dt + A(V_t; \theta)dW_t, \tag{1}$$

where $W_t = (w_{t,1}, \cdots, w_{t,d})^T$ are d independent Brownian motions. $\boldsymbol{b}(V_t; \theta) = (b_1(V_t; \theta), \cdots, b_d(V_t; \theta))^T$ are the drift coefficients, $A(V_t; \theta) = \{a_{i,j}(V_t; \theta)\}_{d \times d}$ are the diffusion coefficients, and θ is the parameter in the coefficients. The methods developed here can be easily extended to time-inhomogeneous processes, in which the drift coefficients and the diffusion coefficients may also depend on the time variable t. In addition, our methods also apply to jump diffusion processes such as

$$dV_t = \boldsymbol{b}(V_t; \theta)dt + A(V_t; \theta)dW_t + dZ_t$$

where Z_t is a compound Poisson process, with the sampling distribution slightly modified to accommodate jumps. An example of the jump process is given in Section 3.2. For clarity, we concentrate on the diffusion process (1).

Without loss of generality, suppose a diffusion process V_t is observed at $t = 0$ and $t = \Delta$. We are interested in generating bridge samples $V_t^{(j)}, j = 1, \cdots, m$, that connect the two observations V_0, V_Δ and follow the target distribution $\pi(V_t) = P(V_t \mid V_0, V_\Delta; \theta)$.

Beskos et al. (2006) proposed to generate continuous sample paths exactly follow-ing the conditional distribution $P(V_t \mid V_0, V_\Delta; \theta)$ by generating "skeleton" samples of Brownian (or Bessel) bridges that are accepted/rejected with a certain probability. This simulation method also provides an unbiased estimate of the transition probabil-ity density $P(V_\Delta \mid V_0, \theta)$. It is shown that this method works well for many processes, though it is limited to reducible processes and sometimes the acceptance rate can be extremely small for certain situations, especially when Δ is large.

Several simulation methods based on the discrete-time approximation of diffusion process have been developed (Brandt and Santa-Clara, 2002; Durham and Gallant, 2002; Elerian et al., 2001; Eraker, 2001; Kloeden and Platen, 1992; Pedersen, 1995; Roberts and Stramer, 2001). In these methods, the time interval $[0, \Delta]$ is divided into M small intervals with equal length $\delta = \Delta/M$ by the intermediate points $t_i = i\delta$, $i = 0, \ldots, M$. Then the continuous diffusion process V_t is approximated by the discrete-time process $(V_{t_0} = V_0, V_{t_1}, \cdots, V_{t_{M-1}}, V_{t_M} = V_\Delta)$ following the distribution

$$P^*(V_{t_1}, \cdots, V_{t_{M-1}} \mid V_{t_0} = V_0, V_{t_M} = V_\Delta; \theta) \propto \prod_{k=1}^{M} P_\delta^*(V_{t_k} \mid V_{t_{k-1}}; \theta), \qquad (2)$$

where

$$P_\delta^*(V_{t_k} \mid V_{t_{k-1}}; \theta) \sim \mathcal{N}\big(V_{t_{k-1}} + \boldsymbol{b}(V_{t_{k-1}}; \theta)\delta, \, A(V_{t_{k-1}}; \theta)A^T(V_{t_{k-1}}; \theta)\delta\big) \qquad (3)$$

is the Euler approximation of the transition probability density function $P_\delta(V_{t_k} \mid V_{t_{k-1}}; \theta)$. $\mathcal{N}(\mu, \Sigma)$ denotes Gaussian distribution with mean μ and covariance Σ, the subscript δ of the transition density denotes the time interval between t_{k-1} and t_k. In the following, we omit the subscript δ for simplicity. When δ is small, $(V_{t_0}, V_{t_1}, \cdots, V_{t_M})$ can usually approximate V_t well. We use $P^*(\cdot)$ to denote the ap-proximated distribution of the discrete-time process $(V_{t_0}, V_{t_1}, \cdots, V_{t_M})$. Other higher order approximation of the true transition probability density, such as the Milstein approximation or Shoji and Ozaki (1998)'s approximation, can also be used in (3).

Generating Monte Carlo samples from the target distribution (2) can be done using Markov chain Monte Carlo method (MCMC) through a transition kernel whose equilibrium distribution is the target distribution (Gilks et al., 1995; Robert and Casella, 1999). However, for most diffusion processes and for large M (in order to

3

achieve approximation accuracy), the mixing rate of MCMC can be very low. To avoid this problem, Roberts and Stramer (2001) and Elerian et al. (2001) proposed to update a block of the bridge sample in one MCMC move, with proposal move developed from a Brownian bridge or an Ornstein-Uhlenbeck bridge, and Beskos et al. (2008) proposed MCMC moves through solving stochastic partial differential equations. Another limitation of MCMC is its difficulty in estimating the normalizing constant of the target distribution, which is $P^*(V_{t_M} \mid V_{t_0}; \theta)$ for the target distribution (2).

In this study, we generate samples that are properly weighted with respect to the target distribution (2) under the framework of sequential Monte Carlo (SMC). In SMC, the bridge samples start at the fixed V_{t_0}, then $V_{t_1}^{(j)}, V_{t_2}^{(j)}, \cdots, V_{t_{M-1}}^{(j)}$ are generated sequentially until the complete bridge samples $(V_{t_0}, V_{t_1}^{(j)}, \cdots, V_{t_{M-1}}^{(j)}, V_{t_M})$ are obtained. The critical issue here is how to utilize the information provided by the end observation V_{t_M} in generating the intermediate states $(V_{t_1}, \cdots, V_{t_{M-1}})$.

The SMC approach for generating diffusion bridges has been proposed and studied (Pedersen, 1995; Durham and Gallant, 2002; Brandt and Santa-Clara, 2002). Their approaches have been shown to be effective in some cases but also fail in other cases. Pedersen (1995) generated the bridge samples through diffusion dynamics without considering the end constraint given by V_{t_M}, thus the generated bridge samples are often far away from V_{t_M} at the end. Durham and Gallant (2002) used linear interpolation to force the bridge samples to move toward V_{t_M}, but ignoring the diffusion dynamics. In this paper we propose an effective resampling scheme in SMC. The resampling is guided by pilots that are generated backward from the end observation V_{t_M} according to the diffusion dynamics. This resampling scheme can be easily combined with other SMC sampling method, including Pedersen (1995)'s sampler and Durham and Gallant (2002)'s sampler.

The rest of the paper is organized as follows. Section 2 introduces the proposed algorithm, with brief backgrounds on SMC, resampling scheme, the optimal resampling priority score for the diffusion bridge problem, and the strategy of generating backward pilots to estimate the optimal resampling priority scores. Section 3 presents two synthetic examples to demonstrate the proposed algorithm. We conclude the article

4

in the last section.

2 Constrained SMC guided by backward resampling

In the following, we use a simpler notation $v_k = V_{t_k}$ to denote the intermediate states. The starting point $v_0 = V_0$ and the end point $v_M = V_\Delta$ are fixed.

2.1 Importance sampling and SMC

In most cases, directly generating samples from high dimensional, constrained distribution (2) is difficult. Based on the importance sampling principle (Marshall, 1956; Robert and Casella, 1999; Liu, 2001), we can draw samples $\boldsymbol{v}^{(j)} \triangleq (v_0, v_1^{(j)}, \cdots, v_{M-1}^{(j)}, v_M)$, $j = 1, \cdots, m$, from a different sampling distribution $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$, and the proper weights of the samples are computed as

$$w^{(j)} = \frac{\prod_{k=1}^{M} P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \theta)}{Q(\boldsymbol{v}^{(j)} \mid v_0, v_M; \theta)}.$$

When the target distribution $P^*(\boldsymbol{v} \mid v_0, v_M; \theta) \propto \prod_{k=1}^{M} P^*(v_k \mid v_{k-1}; \theta)$ is absolutely continuous with respect to the sampling distribution $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$, and $\text{Var}_Q(w) < \infty$,

$$\frac{1}{m} \sum_{j=1}^{m} w^{(j)} = \frac{1}{m} \sum_{j=1}^{m} \frac{\prod_{k=1}^{M} P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \theta)}{Q(\boldsymbol{v}^{(j)} \mid v_0, v_M; \theta)} \tag{4}$$

is an unbiased estimator of the transition probability density $P^*(v_M = V_\Delta \mid v_0 = V_0; \theta)$, which is the normalizing constant of $\prod_{k=1}^{M} P^*(v_k \mid v_{k-1}; \theta)$. The estimator is consistent, that is, it converges to $P^*(v_M = V_\Delta \mid v_0 = V_0; \theta)$ with probability 1 as $m \to \infty$. In addition, for any function $h(\boldsymbol{v})$, if $\text{Var}_Q(w\,h)$ is also finite, then

$$E_{P^*}\big(h(\boldsymbol{v}) \mid v_0, v_M; \theta\big) \simeq \frac{\sum_{j=1}^{m} w^{(j)} h(\boldsymbol{v}^{(j)})}{\sum_{j=1}^{m} w^{(j)}} \tag{5}$$

is a consistent estimator of the expectation of $h(\boldsymbol{v})$ conditional on the given endpoints v_0, v_M. Note that the conditions for obtaining consistent estimators are based

on standard importance sampling principles (Robert and Casella, 1999). It also applies to other target distributions, such as when the underlying process of the target distribution (2) is a jump diffusion process.

The performance of the importance sampling method depends on the choice of the sampling distribution $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$. When the sampling distribution is "perfect", that is

$$Q(\boldsymbol{v} \mid v_0, v_M; \theta) = P^*(\boldsymbol{v} \mid v_0, v_M; \theta),$$

estimator (4) provides the exact value of $P^*(v_M \mid v_0; \theta)$. Although directly generating samples from $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$ and calculating the weights is infeasible in most cases, an efficient sampling distribution $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$ should be close to $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$. For the purpose of efficiency control, the Chi-square divergence between $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$ and $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$, defined as

$$\int \frac{[P^*(\boldsymbol{v} \mid v_0, v_M; \theta)]^2}{Q(\boldsymbol{v} \mid v_0, v_M; \theta)} \, d\boldsymbol{v} - 1 = \mathrm{Var}_Q\left(\overline{w}\right), \tag{6}$$

is often used as a performance measure of the chosen sampling distribution $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$ (Liu, 2001). Here $\overline{w} = w / P^*(v_M \mid v_0; \theta)$ is the *standardized weight*. Assuming the samples generated are independent, the mean square error (MSE) of estimator (5) can be approximated by (Kong et al., 1994; Liu, 1996)

$$\frac{1}{m}\left[E_{P^*}^2(h \mid v_0, v_M; \theta) \, \mathrm{Var}_Q(\overline{w}) + \mathrm{Var}_Q(\overline{w} \, h) - 2E_{P^*}(h \mid v_0, v_M; \theta) \, Cov_Q(\overline{w}, \overline{w} \, h) \right]. \tag{7}$$

Although the minimization of the above MSE depends on the function $h(\boldsymbol{v})$, minimizing $\mathrm{Var}_Q(w)$ (or equivalently $E_{P^*}^2(h \mid v_0, v_M; \theta) \mathrm{Var}_Q(\overline{w})$) is a reasonable and convenient choice in many cases, especially when the expectations of several functions $h(\cdot)$ needs to be evaluated.

Elerian et al. (2001) proposed to use the multivariate normal distribution or the multivariate student-t distribution as $Q(\boldsymbol{v} \mid v_0, v_M; \theta)$. When M is large, it is usually difficult to directly construct a good sampling distribution close to the target distribution $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$ in such a high dimensional space.

In this paper, we draw the samples under the SMC framework, in which the sampling distribution

$$Q(v_1, \cdots, v_{M-1} \mid v_0, v_M; \theta) = \prod_{k=1}^{M-1} r_k(v_k \mid \boldsymbol{v}_{k-1}; \theta)$$

6

is the product of a sequence of conditional distributions. Here $\boldsymbol{v}_k \triangleq (v_0, v_1, \cdots, v_k, v_M)$. At each step k, $k = 1, \cdots, M-1$, we generate $v_k^{(j)}$ from the conditional distribution $r_k(v_k \mid \boldsymbol{v}_{k-1}^{(j)}; \theta)$. More precisely, a straightforward SMC implementation for generating properly weighted bridge samples $\boldsymbol{v}^{(j)}$ can be done through the following algorithm.

Let m be the Monte Carlo sample size. For each j, $j = 1, \ldots, m$,

1. Let $\boldsymbol{v}_0^{(j)} = \{v_0, v_M\}$ and $w_0^{(j)} = 1$.

2. For $k = 1, \cdots, M-1$,

    (a) Draw $v_k^{(j)}$ from distribution $r_k(v_k \mid \boldsymbol{v}_{k-1}^{(j)}; \theta)$. Let $\boldsymbol{v}_k^{(j)} = \{\boldsymbol{v}_{k-1}^{(j)}, v_k\}$.

    (b) Compute the corresponding weight of $\boldsymbol{v}_k^{(j)}$ by

$$w_k^{(j)} = w_{k-1}^{(j)} \frac{P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \theta)}{r_k(v_k^{(j)} \mid \boldsymbol{v}_{k-1}^{(j)}; \theta)}.$$

3. Let $\boldsymbol{v}_M^{(j)} = \boldsymbol{v}_{M-1}^{(j)}$ and $w^{(j)} = w_{M-1}^{(j)} P^*(v_M \mid v_{M-1}^{(j)}; \theta)$.

The advantage of SMC is that we only need to consider the construction of the low dimensional conditional distributions $r_k(v_k \mid \boldsymbol{v}_{k-1}; \theta)$, $k = 1, \cdots, M-1$.

Pedersen (1995) proposed to use the forward equation

$$r_k(v_k \mid \boldsymbol{v}_{k-1}; \theta) = P^*(v_k \mid v_{k-1}; \theta), \tag{8}$$

so that $w_k^{(j)} \equiv 1$ for $k = 1, \cdots, M-1$, and $w^{(j)} = P^*(v_M \mid v_{M-1}^{(j)}; \theta)$. The procedure is essentially generating the forward path $(v_1, \ldots, v_{M-1})$ only, conditioned on the starting point v_0. The path is then forced to connect to the end point v_M at the last step. Because the samples are generated without taking into the account that they have to end at $v_M = V_\Delta$, many of them will have large 'jump's between $v_{M-1}^{(j)}$ and the fixed end v_M. The performance of this simple sampling method is not satisfactory in many cases.

Durham and Gallant (2002) suggested a different sampling distribution

$$r_k(v_k \mid \boldsymbol{v}_{k-1}; \theta) \sim \mathcal{N}\left(v_{k-1} + \frac{v_M - v_{k-1}}{M - k + 1}, \frac{M-k}{M-k+1} A(v_{k-1}; \theta) A^T(v_{k-1}; \theta) \delta\right). \tag{9}$$

This proposal distribution includes a drift term that linearly connects the current position v_{k-1} to the targeted end point v_M, hence forces v_k to move towards v_M as

k increases to M. Stramer and Yan (2006) proved that this sampling distribution is the "perfect" sampling distribution $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$ when both the drift coefficient $\boldsymbol{b}(v_t; \theta)$ and the diffusion coefficient $A(v_t; \theta)$ do not depend on v_t. However, this sampling distribution may not be ideal for some applications, especially when the drift coefficient strongly depends on v_t or when the time interval Δ is large.

For example, consider the Ornstein-Uhlenbeck process

$$dv_t = \theta v_t dt + dw_t. \tag{10}$$

Let $\theta = 0.2$. In Figure 1, we show 100 sample paths (without taking into account the weight) generated from the true conditional distribution $P^*(\boldsymbol{v} \mid v_0 = 0, v_M = 28.3)$ and from Durham and Gallant (2002)'s sampling distribution (9), with time interval $\Delta = 20$ and $M = 400$ intermediate points. It is clear that the sampling distribution (9) does not capture the intrinsic feature of the underlying diffusion bridges, although with a sufficiently large Monte Carlo sample size and proper weighting, the procedure is valid and can be used for inference. Elerian et al. (2001) documented similar findings.
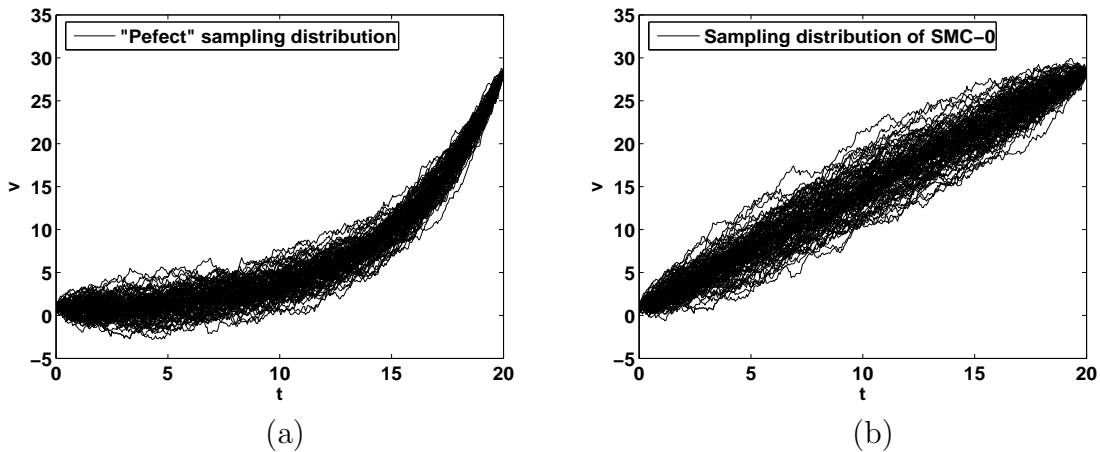

(a) (b)

Figure 1: Illustration of samples of diffusion bridges generated from (a) the true conditional distribution and from (b) Durham and Gallant (2002)'s sampling method for diffusion process (10) with parameter $\theta = 0.2$. $M = 400$ intermediate points are used.

2.2 Resampling and optimal resampling priority score

Resampling (Kong et al., 1994; Liu and Chen, 1998; Liu, 2001) is an important component in SMC to improve efficiency. Its main purpose is to duplicate the 'good' quality partial samples and remove the 'bad' quality partial samples during the sequential build-up of the samples. It provides a way to rejuvenate the samples to improve efficiency in future steps. In our problem the target distribution $P^*(\boldsymbol{v} \mid v_0, v_M; \theta)$ dictates that the samples of $\boldsymbol{v}$ have to connect two fixed points v_0 and v_M, hence imposes a very strong constraint on the sample path. During the sequential build-up, if a partial sample path $\boldsymbol{v}_k$ has moved too far away from the end target v_M and is unlikely to become a valid bridge, then it would be wasting computational resources to continue the build-up to complete the sample, as the complete sample would have very small weight and a negligible contribution to the weighted average used for statistical inferences.

Resampling is done as follows. Suppose we have obtained samples $\{(\boldsymbol{v}_k^{(j)}, w_k^{(j)}), j = 1, \cdots, m\}$ at step k, the resampling step creates a new set of samples $\{(\boldsymbol{v}_k^{new(j_*)}, w_k^{new(j_*)}), j_* = 1, \cdots, m\}$ by drawing samples from the current set $\{\boldsymbol{v}_k^{(j)}, j = 1, \cdots, m\}$ with replacement according to *priority scores* $\{\beta_k^{(j)}, j = 1, \cdots, m\}$, and adjusting the weights accordingly so that (Liu and Chen, 1998)

$$
E\left[\frac{1}{m}\sum_{j_*=1}^{m} w_k^{new(j_*)} h(\boldsymbol{v}_k^{new(j_*)}) \mid \boldsymbol{v}_k^{(j)}, w_k^{(j)}, j = 1, \cdots, m\right] = \frac{1}{m}\sum_{j=1}^{m} w_k^{(j)} h(\boldsymbol{v}_k^{(j)})
$$

for any function $h(\cdot)$. The algorithmic steps are as follows:

1. Assign a priority score $\beta_k^{(j)} > 0$ to each sample $\boldsymbol{v}_k^{(j)}$. Normalize the priority scores so that $\sum_{j=1}^{m} \beta_k^{(j)} = m$.

2. For $j_* = 1, \cdots, m$,

   (a) Randomly sample $\boldsymbol{v}_k^{new(j_*)}$ from the set $\{\boldsymbol{v}_k^{(j)}, j = 1, \cdots, m\}$ with probabilities proportional to the priority scores $\{\beta_k^{(j)}, j = 1, \cdots, m\}$;

   (b) If $\boldsymbol{v}_k^{new(j_*)} = \boldsymbol{v}_k^{(j)}$, then set the new weight associated with $\boldsymbol{v}_k^{new(j_*)}$ to be $w_k^{new(j_*)} = w_k^{(j)}/\beta_k^{(j)}$.

9

3. Return the new set of weighted samples $\{(\boldsymbol{v}_k^{new(j_*)}, w_k^{new(j_*)}), j_* = 1, \cdots, m)\}$.

Here the priority scores are normalized to $\sum_{j=1}^{m} \beta_k^{(j)} = m$ so that the multiplicative constant in the weights does not change.

The priority scores serves as a measure of how 'good' a sample is. In the following we develop the "optimal" priority score for our diffusion bridge problem. Note that the resampling step is to improve the efficiency of future steps, resampling at step $M-1$ is not needed according to Rao-Blackwellization (Liu and Chen, 1998). Hence, we only develop the "optimal" priority score β_k for step $k = 1, \cdots, M-2$. As stated in (6), our goal is to minimize $\mathrm{Var}_Q(w)$.

Suppose at step k ($k \le M-2$), we have obtained sample set $\{(\boldsymbol{v}_k^{(j)}, w_k^{(j)}), j = 1, \cdots, m\}$ from sampling distribution $Q_k(\boldsymbol{v}_k)$ and the corresponding weight

$$w_k = \frac{\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)}{Q_k(\boldsymbol{v}_k)}.$$

Then after a resampling step with priority scores $\beta_k^{(j)}$, the resampled set can be considered as being generated from the sampling distribution $Q_k(\boldsymbol{v}_k)\beta_k$, with new weight

$$w_k^{new} = \frac{\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)}{Q_k(\boldsymbol{v}_k)\beta_k}.$$

Suppose the sampling distribution to generate the future dimensions $(v_{k+1}, \cdots, v_{M-1})$ is $\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)$. In addition, if we do not consider the effect of future resampling steps after step k, then we have

$$\mathrm{Var}_Q(w) = E\left[\frac{\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)}{\beta_k Q_k(\boldsymbol{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)}\right]^2 - \left(P^*(v_M \mid v_0; \theta)\right)^2$$

$$= \int \frac{\left[\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\beta_k Q_k(\boldsymbol{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} dv_1 \cdots dv_{M-1} - \left(P^*(v_M \mid v_0; \theta)\right)^2$$

$$= \int \frac{\left[\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\beta_k Q_k(\boldsymbol{v}_k)} \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} dv_{k+1} \cdots dv_{M-1} dv_1 \cdots dv_k$$

$$- \left(P^*(v_M \mid v_0; \theta)\right)^2.$$

Hence, $\mathrm{Var}_Q(w)$ is minimized when

$$\beta_k Q_k(\boldsymbol{v}_k) \propto \prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta) \left[\int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta) \right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1} \right]^{1/2}.$$

That is,

$$\beta_k \propto w_k \left[\int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta) \right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1} \right]^{1/2}. \tag{11}$$

Especially, when the sampling distribution $\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)$ to generate the future dimensions $(v_{k+1}, \cdots, v_{M-1})$ is "perfect", that is

$$\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta) = P^*(v_{k+1}, \cdots, v_{M-1} \mid v_k, v_M; \theta) = \frac{\prod_{s=k+1}^{M} P_s^*(v_s \mid v_{s-1}, v_M; \theta)}{P^*(v_M \mid v_k; \theta)},$$

the corresponding "optimal" resampling priority score (11) becomes

$$\beta_k = w_k P(v_M \mid v_k; \theta).$$

In this case, the priority score is proportional to the transition probability from the current position v_k to the fixed end point v_M, though evaluation of this quantity has the same difficulty as our original problem.

The optimal sampling distribution of Liu and Chen (1998) provides another interpretation of this resampling priority score. Under our setting, the optimal sampling distribution at time k would be proportional to

$$P^*(v_k \mid v_{k-1}, v_M) \propto P^*(v_k \mid v_{k-1}; \theta) P^*(v_M \mid v_k; \theta)$$

One way to draw sample from this optimal distribution is to use the first term to sample and the second term as the resampling priority score.

2.3 Resampling guided by backward pilots

A difficulty in obtaining the optimal priority score assignment (11) is that the value of integration

$$f_k(v_k; \theta) \triangleq \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta) \right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1}, \quad k = 1, \cdots, M-2, \tag{12}$$

is unknown. Here we use a separate SMC procedure to generate pilot samples $(u_k, u_{k+1}, \ldots, u_{M-1}, u_M = v_M)$ backward from the fixed point u_M and obtain an estimate of the function.

Specifically, at step k, $k = M - 1, \cdots, 1$, we generate pilots $\boldsymbol{u}_k^{(j)} \triangleq (u_M^{(j)} = v_M, u_{M-1}^{(j)}, \cdots, u_k^{(j)})$, $j = 1, \cdots, m^*$ that are properly weighted by $a_k^{(j)}$ with respect to the distribution proportional to

$$\frac{\left[\prod_{s=k+1}^{M} P^*(u_s \mid u_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(u_s \mid u_{s-1}, u_M; \theta)},$$

then $f_k(v_k; \theta)$ can be estimated through the weighted pilots $\{(\boldsymbol{u}_k^{(j)}, a_k^{(j)}), j = 1, \cdots, m^*\}$. The specific algorithmic steps are as follows

1. For $k = M$, let $a_M^{(j)} = 1$, $u_M^{(j)} = v_M$, $j = 1, \cdots, m^*$.

2. For $k = M - 1, M - 2, \cdots, 1$, do, for each $j = 1, \ldots, m^*$,

   (a) Generate $u_k^{(j)}$ from a sampling distribution $g_k(u_k \mid u_{k+1}^{(j)}; \theta)$. The choice of $g_k(u_k \mid u_{k+1}^{(j)}; \theta)$ will be discussed in the following Remark 2.

   (b) Calculate the corresponding weight of $\boldsymbol{u}_k^{(j)}$ by

   $$a_k^{(j)} = \begin{cases} a_M^{(j)} \dfrac{\left[P^*(u_M^{(j)} \mid u_{M-1}^{(j)}; \theta)\right]^2}{g_{M-1}(u_{M-1}^{(j)} \mid u_M^{(j)}; \theta)} & \text{if } k = M - 1, \\[4ex] a_{k+1}^{(j)} \dfrac{\left[P^*(u_{k+1}^{(j)} \mid u_k^{(j)}; \theta)\right]^2}{r_{k+1}(u_{k+1}^{(j)} \mid u_k^{(j)}, u_M^{(j)}; \theta) g_k(u_k^{(j)} \mid u_{k+1}^{(j)}; \theta)} & \text{if } k \leq M - 2. \end{cases}$$

   (c) If $k \leq M - 2$, estimate the function $f_k(v_k; \theta)$ in (12) using $\{(u_k^{(j)}, a_k^{(j)}), j = 1, \cdots, m^*\}$, through any kind of density estimator such as

   $$\widehat{f}_k(v; \theta) = \sum_{j=1}^{m^*} K_h(v - u_k^{(j)}) a_k^{(j)}$$

   where $K_h(\cdot)$ is a kernel function with bandwidth h; or a histogram estimator, with partition $\mathbb{D}_{k,1} \bigcup \mathbb{D}_{k,2} \ldots \bigcup \mathbb{D}_{k,n_k}$ of the state space of v_k and

   $$\widehat{f}_k(v; \theta) = \sum_{l=1}^{n_k} \widehat{f}_{k,l} \mathbb{I}(v \in \mathbb{D}_{k,l}),$$

12

where

$$\widehat{f}_{k,l} = \frac{1}{m^* \, |\mathbb{D}_{k,l}|} \sum_{j=1}^{m^*} a_k^{(j)} \mathbb{I}(u_k^{(j)} \in \mathbb{D}_{k,l}),$$

and $|\mathbb{D}_{k,l}|$ is the volume of subspace $\mathbb{D}_{k,l}$, $\mathbb{I}(\cdot)$ is the indicator function.

(d) (Optional) Perform resampling to the backward pilots $\{(\boldsymbol{u}_k^{(j)}, a_k^{(j)}), j = 1, \cdots, m^*\}$ with priority score proportional to $a_k^{(j)}$ if necessary (Liu and Chen, 1998).

Figure 2 depicts the idea. Two forward (partial) bridge samples (a) and (b) have been generated up to time $k = 100$. The backward pilots samples up to $k = 100$ are shown and the histogram estimate of f_k at $k = 100$ based on these backward pilot samples are also shown (vertically) on the right side of the figure. It is seen that path (b) as a partial sample of the bridge is a better sample because it has higher probability to connect to the fixed end than path (a).
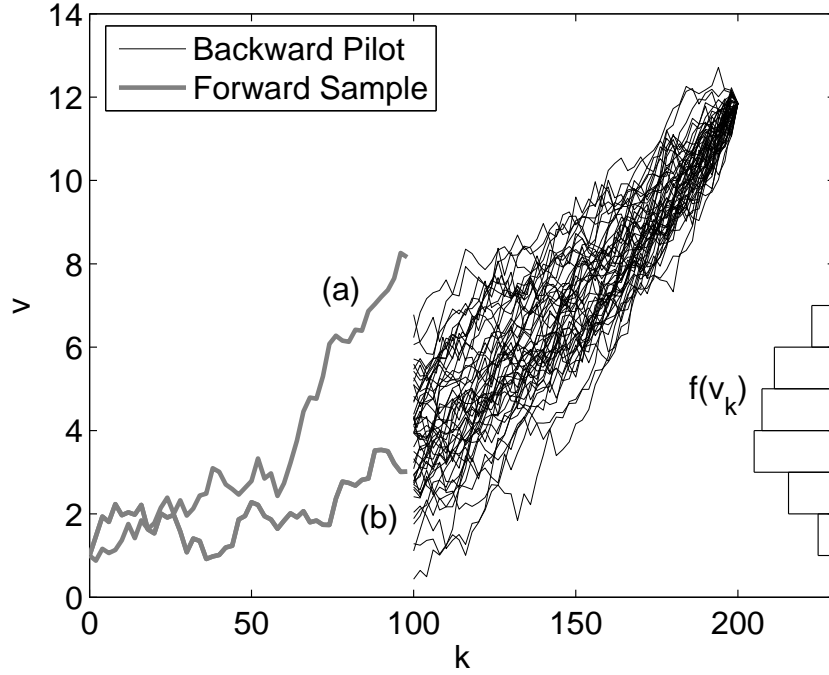

Figure 2: Illustration of using the backward pilots to obtain the resampling priority score.

Remarks:

13

1. The backward pilots $\boldsymbol{u}_k^{(j)}$, $j = 1, \cdots, m^*$, only need to be generated once from $k = M - 1$ to $k = 1$ before we generate the diffusion bridge samples. The function $f_k(\cdot)$ for all $k = 1, \cdots, M - 2$ are estimated in this process. Hence the extra computational burden for calculating $f_k(\cdot)$ is limited. In addition, this stage serves as a general guidance for resampling. A very accurate estimation of the function $f_k(\cdot)$ is not necessary because resampling relies more on the global picture. Accurate details of $f_k(\cdot)$ will not improve the resampling performance significantly. Hence the number of backward pilots m^* needed does not necessarily have to be large.

2. We choose the sampling distribution $g_k(u_k \mid u_{k+1}^{(j)}; \theta)$ in step 2(a) to be approximately proportional to $P^*(u_{k+1}^{(j)} \mid u_k; \theta)$. If we consider the forward sampling distribution $\prod_{s=k+1}^{M-1} r_s(u_s \mid u_{s-1}, u_M; \theta)$ as an approximation of $\prod_{s=k+1}^{M} P^*(u_s \mid u_{s-1}; \theta)$, then $f_k(v_k; \theta)$ can be approximated by $\int \prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta) \, dv_{k+1} \cdots dv_{M-1}$. Hence, $g_k(u_s \mid u_{k+1}; \theta) \propto P^*(u_{k+1} \mid u_k; \theta)$ is a reasonable choice of the sampling distribution for estimating $f_k(v_k; \theta)$. In the Euler approximation, we have

$$u_{k+1} = u_k + \boldsymbol{b}(u_k; \theta)\delta + A(u_k; \theta)\varepsilon_k,$$

where ε_k is d-dimensional Gaussian distribution $\mathcal{N}(0, \delta \, I_d)$, I_d is the identity matrix. To approximate $P^*(u_{k+1}^{(j)} \mid u_k; \theta)$, we apply Taylor expansion to $\boldsymbol{b}(u_k; \theta)$ at point $u_k^{*(j)} = u_{k+1}^{(j)} - \boldsymbol{b}(u_{k+1}^{(j)}; \theta)\delta$ and assume $A(u_k; \theta)$ is constant $A(u_k^{*(j)}; \theta)$; then we have

$$u_{k+1}^{(j)} = \left(I + H(u_k^{*(j)}; \theta)\delta\right) u_k + \boldsymbol{b}(u_k^{*(j)}; \theta)\delta - H(u_k^{*(j)}; \theta)u_k^{*(j)}\delta + A(u_k^{*(j)}; \theta)\varepsilon_k,$$

where $H(u; \theta)$ is the Jacobi matrix of the drift coefficient $\boldsymbol{b}(u; \theta)$. Hence, to make $g_k(u_k \mid u_{k+1}^{(j)}; \theta)$ approximately proportional to $P^*(u_{k+1} \mid u_k, \theta)$, we can let $g_k(u_k \mid u_{k+1}^{(j)}; \theta) \sim \mathcal{N}\left(\mu_k^{(j)}, \Sigma_k^{(j)}\right)$, where

$$\mu_k^{(j)} = \left(\Sigma_k^{*(j)}\right)^{-1} \left(u_{k+1}^{(j)} - \boldsymbol{b}(u_k^{*(j)}, \theta)\delta + H(u_k^{*(j)}; \theta)u_k^{*(j)}\delta\right),$$

$$\Sigma_k^{(j)} = \left(\Sigma_k^{*(j)}\right)^{-1} A(u_k^{*(j)}; \theta)A^T(u_k^{*(j)}; \theta) \left(\Sigma_k^{*(j)}\right)^{-T} \delta,$$

$$\Sigma_k^{*(j)} = I + H(u_k^{*(j)}; \theta)\delta.$$

3. When generating the bridge samples $\boldsymbol{v}^{(j)}$, $j = 1, \cdots, m$, using the algorithm in Section 2.1, the "optimal" resampling priority scores at step k should be set to

$$\beta_k^{(j)} \propto w_k^{(j)} \left[\widehat{f}_k(v_k^{(j)}; \theta) \right]^{-1/2}. \tag{13}$$

4. In general, the backward pilot scheme can be used to estimate any function of v_k in the form of

$$\int \zeta(v_k, v_{k+1}, \cdots, v_{M-1}, v_M) \, dv_{k+1} \cdots dv_{M-1},$$

including the transition probability (as a function of v_k)

$$P^*(v_M \mid v_k; \theta) = \int \prod_{s=k}^{M-1} P^*(v_{s+1} \mid v_s; \theta) \, dv_{k+1} \cdots dv_{M-1}.$$

5. In step 2(c) of the algorithm, although the function $\widehat{f}_k(\cdot)$ estimated using kernel estimator often has good properties, it can be computationally expensive to evaluate when the Monte Carlo sample size m^* is large. Our experience has shown that the histogram estimator is sufficient in most cases.

3   Examples

3.1   Example 1

Beskos et al. (2006) considered diffusion process v_t characterized by stochastic diffusion equation (SDE)

$$dv_t = sin(v_t - \theta)dt + dw_t, \tag{14}$$

where $sin(v_t - \theta)$ is the drift coefficient, θ is the parameter, w_t is a Brownian motion. This diffusion process actually exhibits certain jump phenomena. It is observed that when v_t falls in the interval $(\theta + 2k\pi, \theta + 2(k+1)\pi)$, $k \in \mathbb{Z}$, the drift function will pull the process quickly towards $\theta + \pi + 2k\pi$. Figure 3 shows a realization of the process with parameter $\theta = \pi$.
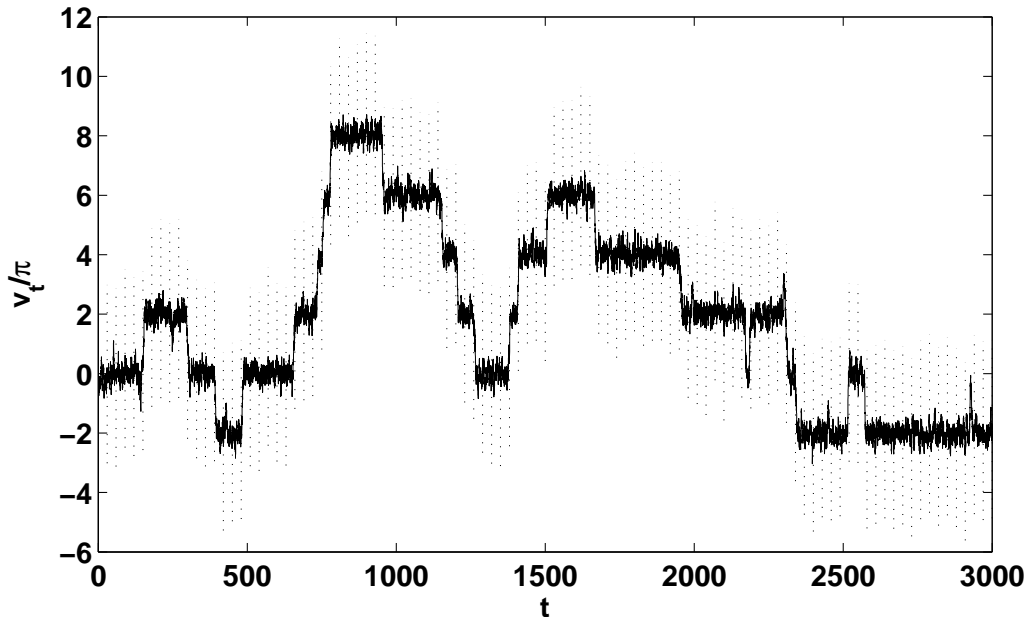
Figure 3: A realization of the process with sine drift coefficients following equation (14), with $\theta = \pi$. The vertical lines present observations with time interval $\Delta = 30$.

3.1.1 Estimation of the transition density with parameter $\theta = \pi$

We first consider the generation of the diffusion bridges with two fixed end points at V_0 and V_{30}, using $M = 400$ intermediate time points. To simplify the notation, we use SMC-0 to denote Durham and Gallant (2002)'s sampler (9) without resampling, and use SMC-1 to denote Durham and Gallant (2002)'s sampler (9) with resampling steps according to the "optimal" priority score (13). In SMC-1, we use the histogram estimator for the function f_k with partition $\bigcup_l \mathcal{D}_l \triangleq \bigcup_l [\frac{\pi}{3} l + \theta - \frac{\pi}{6}, \frac{\pi}{3} l + \theta + \frac{\pi}{6})$, $l \in \mathbb{Z}$, and $m^* = 300$ backward pilots. The resampling step is performed every 20 steps when generating the bridge samples.

Figure 4 shows 100 samples of the diffusion bridge connecting four consecutive observations $(V_0, V_{30}, V_{60}, V_{90}) = (0, 1.49, -5.91, -1.17)$ using different sampling methods. The samples under "perfect" sampling (Figure 4(a)) are obtained by first generating 20,000 samples from SMC-0 then choosing 100 samples among them with probability proportional to their corresponding final weights $w^{(j)}$. The samples of SMC-0 (Figure 4(b)) and SMC-1 (Figure 4(c)) are obtained by generating 100 sam-

ples using the corresponding methods. Here we do not weight the samples, just to show the properties of the generated samples and the corresponding trial distributions. Clearly the sampling distribution of SMC-1 is much closer to the "perfect" sampling distribution and captures the "jump" behavior of the diffusion bridge. It is also interesting to note that the timing of the jumps occurs almost uniformly within the time interval. As there is no information other than the two end-points, the uniformity of the jump timing is quite expected.

As in Durham and Gallant (2002), we use

$$
E_Q \left[\log \left(\frac{1}{m} \sum_{j=1}^{m} w^{(j)} \right) - \log P^*(v_M \mid v_0; \theta) \right]^2
$$

as the measurement of the efficiency for different sampling method. In fact, when $\boldsymbol{v}^{(j)}$, $j = 1, \cdots, m$, are generated independently, we have

$$
\begin{aligned}
& E_Q \left[\log \left(\frac{1}{m} \sum_{j=1}^{m} w^{(j)} \right) - \log P^*(v_M \mid v_0; \theta) \right]^2 \\
& \approx \quad E_Q \left[\frac{\frac{1}{m} \sum_{j=1}^{m} w^{(j)}}{P^*(v_M \mid v_0; \theta)} - 1 \right]^2 = \frac{1}{m} \mathrm{Var}_Q \left[\frac{w^{(j)}}{P^*(v_M \mid v_0; \theta)} \right].
\end{aligned}
$$

Hence, this measurement can be treated as an approximation of measurement (6) divided by the sample size m. For fair comparison, we adjust the sample size m so that different sampling methods take about the same computational time.

Fixing parameter θ at π, for each pair of end-points V_0 and V_{30}, we repeat the estimation 100 times independently and calculate

$$
\mathrm{RMSE}(V_0, V_{30}) \quad = \quad \left[\frac{1}{100} \sum_{i=1}^{100} \left(\log \widehat{P}^{*(i)}(V_{30} \mid V_0, \theta = \pi) - \log P(V_{30} \mid V_0, \theta = \pi) \right)^2 \right]^{1/2},
$$

as the performance measurement, where $\widehat{P}^{*(i)}(V_{30} \mid V_0, \theta = \pi)$ is the i-th independent estimate of $P^*(V_{30} \mid V_0, \theta = \pi)$. The "true" value of $\log P(V_{30} \mid V_0; \theta = \pi)$ is obtained by using Beskos et al. (2006)'s exact sampling method with $m = 10,000,000$ Monte Carlo samples. With roughly equal computational time, we used $m = 3,500$ samples for SMC-0 and $m = 1,000$ samples for SMC-1. Note that the sampling distributions of SMC-0 does not depend on the parameter and the samples can be linearly transformed
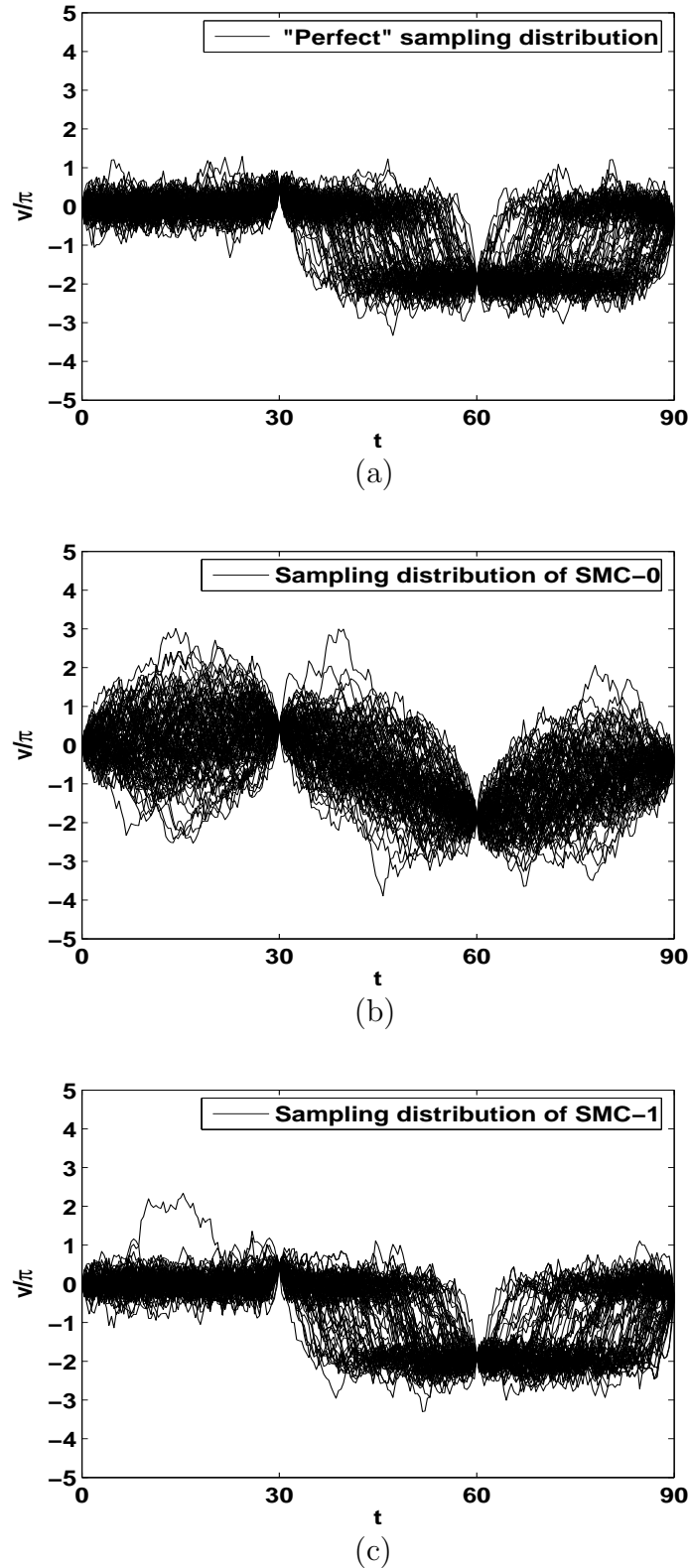
Figure 4: Illustration of bridge samples of the sine drift process generated using different sampling methods with $M = 400$ intermediate points between two consecutive observations. The parameter is $\theta = \pi$ and the observations are $(V_0, V_{30}, V_{60}, V_{90}) = (0, 1.49, -5.91, -1.17)$. (a): The "perfect" sampling distribution; (b): Durham and Gallant (2002)'s sampling method (SMC-0); (c): Durham and Gallant (2002)'s sampling method with resampling steps (SMC-1). In SMC-1, the resampling step is performed every 20 steps when generating the bridge samples. $m^* = 300$ backward pilots are generated to estimate the resampling priority scores.

to meet different fixed end-points, though the weight calculation depends on θ and the end-points. Hence the computational time for SMC-0 only involves the time for evaluating $\prod_{k=1}^{M} P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \theta = \pi)$.

Table 1 reports the rate of RMSE(V_0, V_{30}) of SMC-1 over RMSE(V_0, V_{30}) of SMC-0. it is seen that for most of the (V_0, V_{30}) pairs when V_{30} is not too far away from V_0, SMC-1 has better estimation accuracy. However, when $|V_{30} - V_0| > 4\pi$, SMC-0 outperforms SMC-1. In this case, the process is required to jump more than two levels within a short time period $\Delta = 30$. Simulation shows that such cases happen only 0.1% of the times. Table 2 and Table 3 report RMSE(V_0, V_{30}) of SMC-0 and SMC-1, respectively. It shows that RMSE(V_0, V_{30}) of SMC-0 decreases slowly as $|V_{30} - V_0|$ increases. However, RMSE(V_0, V_{30}) of SMC-1 maintains a relative small value when V_0 and V_{30} are close, but increases fast with the distance between V_0 and V_{30} when $|V_{30} - V_0| > 4\pi$. It seems that when there are several one-directional jumps between the two end-points in short time interval, the most likely paths are those close to the straight line between the two end-points. In this case, the sampling distribution employed by Durham and Gallant (2002) (SMC-0) guides the sampling more forcefully to reach the far-away target endpoint. On the other hand, the resampling approach (SMC-1) may require larger MC sample size of the backward pilots to obtain "good" resampling priority scores, as the backward pilots are generated without considering the location of v_0. The log transition probability estimated by SMC-1 with $m = 1,000$ samples is plotted in Figure 5.

Figure 6 shows the sampled diffusion bridges for selected pairs of (V_0, V_{30}) using SMC-1 method. It is seen that the process is stable around $0, 2\pi, 4\pi, \ldots$ and is transient around $\pi, 3\pi, etc.$, and exhibits jump behavior once getting into the transition zone, spending very short time inside the transit zone. Hence, the process tends to stay in the stable zone if not required to move to another zone (Panel (a)). If the end point is in the transit zone, then the process tends to stay in the stable zone as long as it can, then move into the transit zone at the end to meet the ending requirement (Panels (b), (d) and (f)). Panel (b) also shows that the process is also able to jumps to a different stable zone and then come back to meet the boundary requirement at the end. If the starting point and the ending point are both in stable

19

$\dfrac{\mathrm{RMSE_{SMC\text{-}1}}\left(V_0,V_{30}\right)}{\mathrm{RMSE_{SMC\text{-}0}}\left(V_0,V_{30}\right)}$	V_0										
	-1.0π	-0.8π	-0.6π	-0.4π	-0.2π	0	0.2π	0.4π	0.6π	0.8π	1.0π
$V_{30}=-4.8\pi$	0.48	0.74	0.92	**1.20**	**1.48**	**1.72**	**1.66**	**1.69**	**2.06**	**1.91**	**3.09**
$V_{30}=-4.2\pi$	0.59	0.69	0.83	**1.16**	**1.45**	**1.56**	**1.52**	**2.14**	**1.82**	**1.81**	**2.39**
$V_{30}=-3.6\pi$	0.47	0.62	0.81	0.96	**1.17**	**1.38**	**1.27**	**1.39**	**1.68**	**1.39**	**1.76**
$V_{30}=-3.0\pi$	0.39	0.34	0.43	0.45	0.56	0.49	0.59	0.56	0.58	0.47	0.50
$V_{30}=-2.4\pi$	0.34	0.32	0.43	0.51	0.44	0.49	0.59	0.53	0.43	0.65	0.47
$V_{30}=-1.8\pi$	0.46	0.40	0.48	0.45	0.56	0.57	0.53	0.57	0.42	0.42	0.55
$V_{30}=-1.2\pi$	0.35	0.38	0.46	0.50	0.45	0.40	0.50	0.50	0.33	0.45	0.41
$V_{30}=-0.6\pi$	0.32	0.25	0.36	0.42	0.32	0.37	0.30	0.37	0.27	0.26	0.28
$V_{30}=0.0\pi$	0.32	0.32	0.47	0.41	0.36	0.34	0.37	0.40	0.27	0.30	0.30
$V_{30}=0.6\pi$	0.61	0.28	0.32	0.33	0.41	0.27	0.34	0.25	0.30	0.29	0.45
$V_{30}=1.2\pi$	0.40	0.40	0.39	0.43	0.39	0.55	0.55	0.63	0.49	0.52	0.30
$V_{30}=1.8\pi$	0.45	0.54	0.51	0.55	0.63	0.45	0.55	0.37	0.40	0.38	0.37
$V_{30}=2.4\pi$	0.49	0.54	0.53	0.49	0.52	0.55	0.49	0.54	0.44	0.41	0.28
$V_{30}=3.0\pi$	0.61	0.58	0.50	0.41	0.57	0.49	0.45	0.64	0.49	0.40	0.43
$V_{30}=3.6\pi$	**1.54**	**1.53**	**1.30**	**1.34**	**1.38**	**1.15**	**1.41**	**1.12**	0.87	0.56	0.58
$V_{30}=4.2\pi$	**2.01**	**1.77**	**1.62**	**2.04**	**1.29**	**1.67**	**1.03**	**1.06**	0.86	0.71	0.51
$V_{30}=4.8\pi$	**2.66**	**2.55**	**1.67**	**1.83**	**1.51**	**1.54**	**1.60**	**1.51**	0.91	0.55	0.56

Table 1: The ratio of $\mathrm{RMSE}(V_0, V_{30})$ between SMC-1 and SMC-0 for estimating $\log P(V_{30} \mid V_0; \theta = \pi)$ for the diffusion process with sine drift. The samples sizes ($m = 3,500$ for SMC-0 and $m = 1,000$ for SMC-1) are controlled so the two methods take similar computational time.

$\mathrm{RMSE}_{\mathrm{SMC\text{-}0}}\!\left(V_0, V_{30}\right)$	V_0										
	-1.0π	-0.8π	-0.6π	-0.4π	-0.2π	0	0.2π	0.4π	0.6π	0.8π	1.0π
$V_{30} = -4.8\pi$	0.27	0.26	0.23	0.25	0.20	0.24	0.22	0.21	0.17	0.19	0.17
$V_{30} = -4.2\pi$	0.23	0.23	0.25	0.19	0.22	0.20	0.19	0.14	0.17	0.19	0.20
$V_{30} = -3.6\pi$	0.32	0.28	0.27	0.26	0.22	0.23	0.22	0.21	0.20	0.23	0.20
$V_{30} = -3.0\pi$	0.40	0.44	0.33	0.30	0.27	0.29	0.24	0.26	0.32	0.29	0.31
$V_{30} = -2.4\pi$	0.33	0.36	0.29	0.26	0.25	0.26	0.21	0.31	0.29	0.25	0.26
$V_{30} = -1.8\pi$	0.41	0.31	0.36	0.30	0.24	0.21	0.23	0.27	0.29	0.31	0.26
$V_{30} = -1.2\pi$	0.47	0.42	0.45	0.42	0.36	0.33	0.31	0.33	0.44	0.34	0.42
$V_{30} = -0.6\pi$	0.47	0.43	0.44	0.38	0.34	0.30	0.37	0.33	0.41	0.42	0.42
$V_{30} = 0.0\pi$	0.35	0.32	0.33	0.30	0.27	0.26	0.27	0.36	0.33	0.39	0.37
$V_{30} = 0.6\pi$	0.35	0.41	0.45	0.32	0.35	0.33	0.28	0.39	0.41	0.38	0.49
$V_{30} = 1.2\pi$	0.40	0.39	0.38	0.35	0.35	0.33	0.43	0.30	0.38	0.39	0.47
$V_{30} = 1.8\pi$	0.27	0.27	0.22	0.26	0.22	0.27	0.27	0.30	0.34	0.35	0.38
$V_{30} = 2.4\pi$	0.26	0.32	0.24	0.26	0.23	0.22	0.26	0.24	0.34	0.38	0.39
$V_{30} = 3.0\pi$	0.25	0.27	0.27	0.29	0.29	0.25	0.32	0.28	0.31	0.44	0.53
$V_{30} = 3.6\pi$	0.26	0.22	0.21	0.24	0.24	0.25	0.21	0.24	0.26	0.34	0.27
$V_{30} = 4.2\pi$	0.20	0.18	0.17	0.17	0.20	0.18	0.26	0.22	0.22	0.24	0.25
$V_{30} = 4.8\pi$	0.18	0.16	0.21	0.19	0.18	0.24	0.20	0.20	0.24	0.26	0.28

Table 2: $\mathrm{RMSE}(V_0, V_{30})$ using SMC-0 when estimating $\log P(V_{30} \mid V_0; \theta = \pi)$ of the diffusion process with sine drift, with sample size $m = 3,500$.

$\mathrm{RMSE}_{\text{SMC-1}}(V_0, V_{30})$	V_0										
	-1.0π	-0.8π	-0.6π	-0.4π	-0.2π	0	0.2π	0.4π	0.6π	0.8π	1.0π
$V_{30} = -4.8\pi$	0.13	0.19	0.21	0.30	0.29	0.42	0.36	0.36	0.35	0.37	0.53
$V_{30} = -4.2\pi$	0.14	0.16	0.21	0.22	0.31	0.31	0.28	0.30	0.30	0.34	0.48
$V_{30} = -3.6\pi$	0.15	0.17	0.22	0.25	0.26	0.32	0.28	0.29	0.33	0.32	0.36
$V_{30} = -3.0\pi$	0.15	0.15	0.14	0.14	0.15	0.14	0.14	0.14	0.19	0.14	0.15
$V_{30} = -2.4\pi$	0.11	0.12	0.12	0.13	0.11	0.13	0.12	0.16	0.12	0.17	0.12
$V_{30} = -1.8\pi$	0.19	0.12	0.17	0.13	0.13	0.12	0.12	0.15	0.12	0.13	0.14
$V_{30} = -1.2\pi$	0.17	0.16	0.21	0.21	0.16	0.13	0.15	0.17	0.14	0.15	0.17
$V_{30} = -0.6\pi$	0.15	0.10	0.16	0.16	0.11	0.11	0.11	0.12	0.11	0.11	0.12
$V_{30} = 0.0\pi$	0.11	0.10	0.15	0.12	0.10	0.09	0.10	0.15	0.09	0.12	0.11
$V_{30} = 0.6\pi$	0.21	0.12	0.14	0.11	0.14	0.09	0.09	0.10	0.12	0.11	0.22
$V_{30} = 1.2\pi$	0.16	0.15	0.15	0.15	0.14	0.18	0.24	0.19	0.18	0.20	0.14
$V_{30} = 1.8\pi$	0.12	0.15	0.11	0.14	0.14	0.12	0.15	0.11	0.14	0.13	0.14
$V_{30} = 2.4\pi$	0.13	0.17	0.13	0.13	0.12	0.12	0.13	0.13	0.15	0.16	0.11
$V_{30} = 3.0\pi$	0.15	0.16	0.14	0.12	0.17	0.13	0.14	0.18	0.15	0.17	0.23
$V_{30} = 3.6\pi$	0.40	0.33	0.27	0.31	0.33	0.29	0.29	0.27	0.23	0.19	0.15
$V_{30} = 4.2\pi$	0.41	0.32	0.28	0.35	0.26	0.30	0.26	0.23	0.19	0.17	0.13
$V_{30} = 4.8\pi$	0.48	0.41	0.35	0.35	0.28	0.36	0.32	0.31	0.22	0.15	0.16

Table 3: $\mathrm{RMSE}(V_0, V_{30})$ using SMC-1 when estimating $\log P(V_{30} \mid V_0; \theta = \pi)$ of the diffusion process with sine drift, with sample size $m = 1,000$.
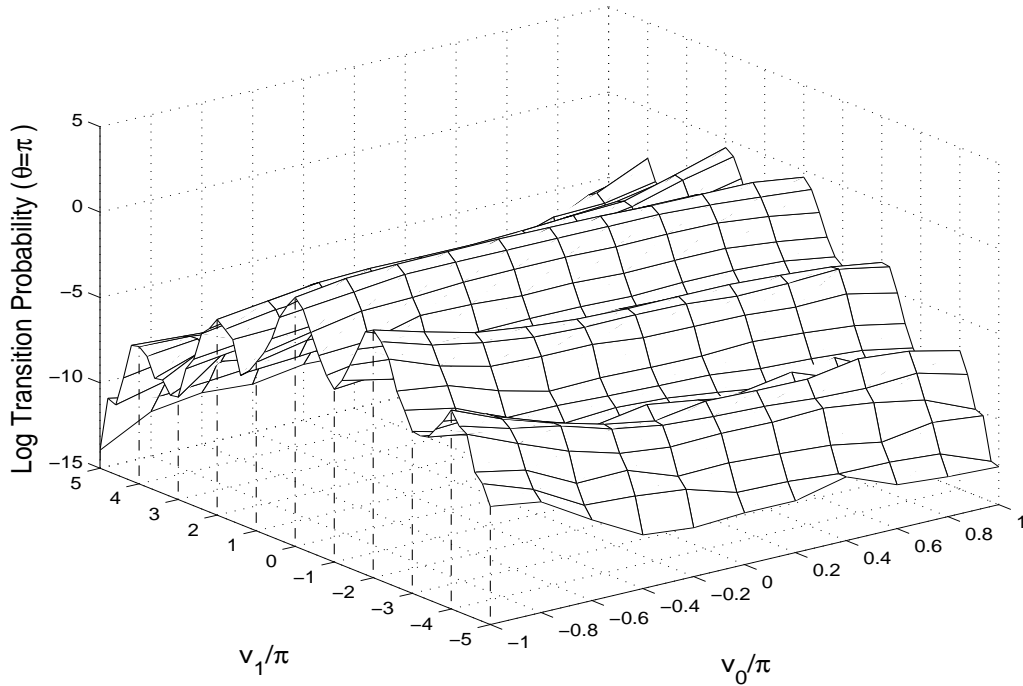
Figure 5: This figure shows the log transition probability of the sine drift process estimated using SMC-1. The sample size is $m = 1,000$.

but different zones, then the jump will occur within the time period and the timing is almost uniform, except at the beginning and the end (Panels (c) and (e)).

3.1.2 Likelihood function estimation

The ability to estimate the transition probability also allows us to estimate the likelihood function. Given observations observed at discrete time $V_{t_0}, V_{t_1}, \ldots, V_{t_n}$, the log-likelihood function is

$$L(\theta) = \sum_{i=1}^{n} L_i(\theta) = \sum_{i=1}^{n} \log P(V_{t_i} \mid V_{t_{i-1}}; \theta).$$

In the following we investigate the performance of the proposed method for likelihood function estimation.

We simulate 100 paths of the process in (14) with $\theta = \pi$, each with $n = 101$ observations and time interval $\Delta = 30$ between two observations. Hence the observations are observed at $t = 0, 30, 60, \ldots, 3000$. The paths are simulated using Euler
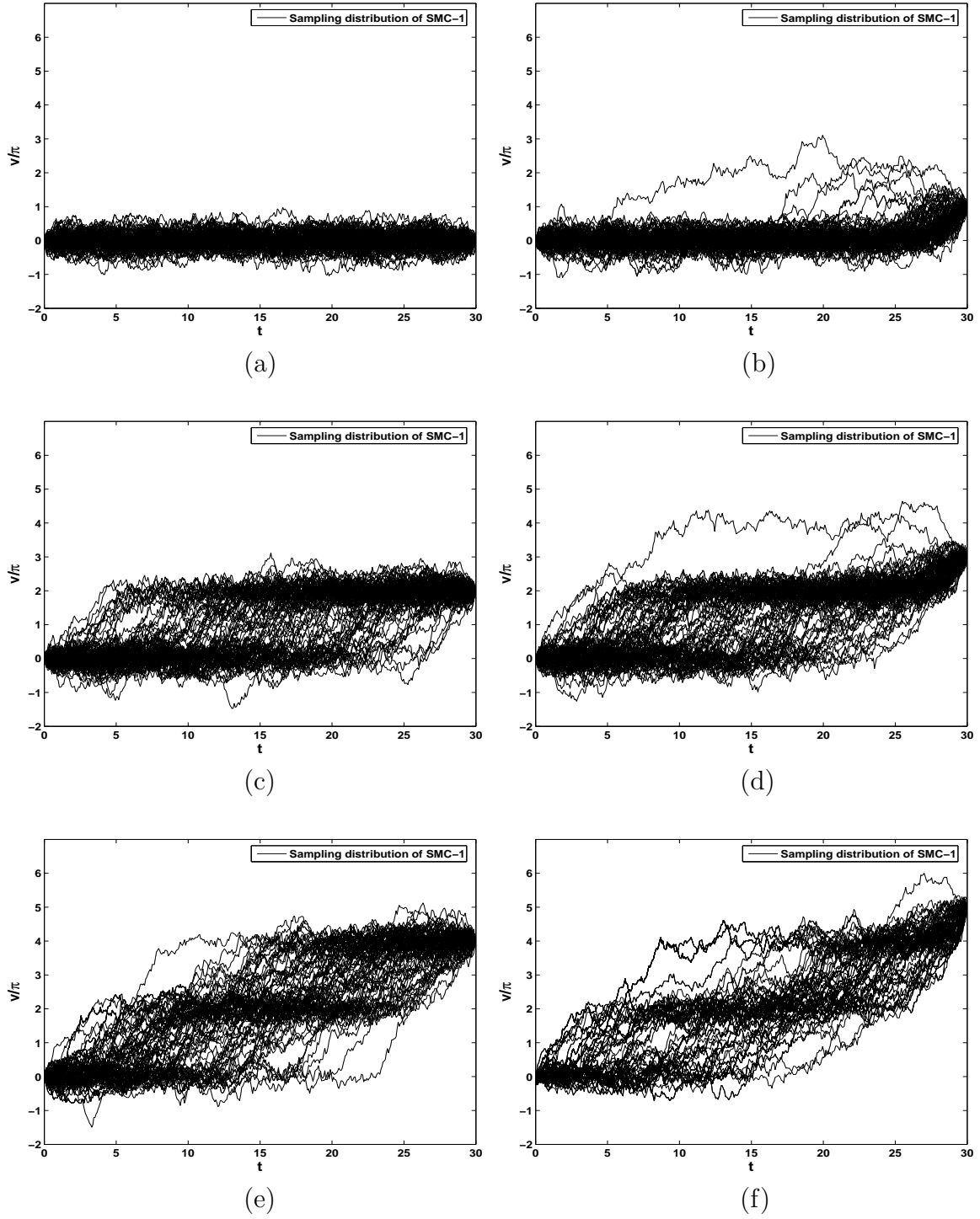
23

Figure 6: Illustration of bridge samples of the sine drift process generated by SMC-1 for $\theta = \pi$, $V_0 = 0$ and different values of V_{30}. The resampling step is performed every 20 steps when generating the bridge samples. $m^* = 300$ backward pilots are generated to estimate the resampling priority scores.

24

approximation with a very small time step ($\Delta/10,000$). We compare the efficiency of the exact sampling method proposed by Beskos et al. (2006), SMC-0 and SMC-1.

The measurement of efficiency we use is

$$\text{RMSE}(\theta) \quad = \quad \left[\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{L}_i(\theta) - L_i(\theta) \right)^2 \right]^{1/2}.$$

Again, the "true" value of $L_i(\theta)$ is obtained by using the exact sampling method with $m = 10,000,000$ Monte Carlo sample size.

We report the average $\text{RMSE}(\theta)$ of the 100 simulated paths for different methods in Table 4. For each method, its Monte Carlo sample size m is chosen so that the methods take approximately the same CPU time. From the table we can see that SMC-1 performs the best for estimating $L(\theta)$ of all θ (the true parameter is at $\theta = \pi$). Although the exact sampling method Beskos et al. (2006) produces unbiased estimators of the transition density hence the likelihood function, its performance is not satisfactory, due to the high rejection rate in the sampling process.

In Figure 7 we plot the log-likelihood function of θ in $[0.76\pi, 1.24\pi]$ with a grid of every 0.02π, based on one set of observations at $t = 0, 30, 60, \ldots, 3000$. The solid line is the estimated log-likelihood function using SMC-1 with 1000 samples. For comparison, the dashed line plots the "true" log-likelihood function using SMC-0 with 100,000 samples. The diamond ($\hat{\theta} = 1.04\pi$) and the circle ($\hat{\theta} = 1.06\pi$) shown on the plot are the MLE using the estimated (SMC-1) and the 'true' log-likelihood functions, respectively. We can see that the estimated log-likelihood function is close to the true one, but is not smooth at places, due to randomness of the Monte Carlo samples. However, the MLE is quite close to the "true" one. It is possible to combine the proposed method with the smooth particle filter (Pitt, 2002) to obtain an improved estimation of a smooth likelihood function. Further research on this direction is needed.
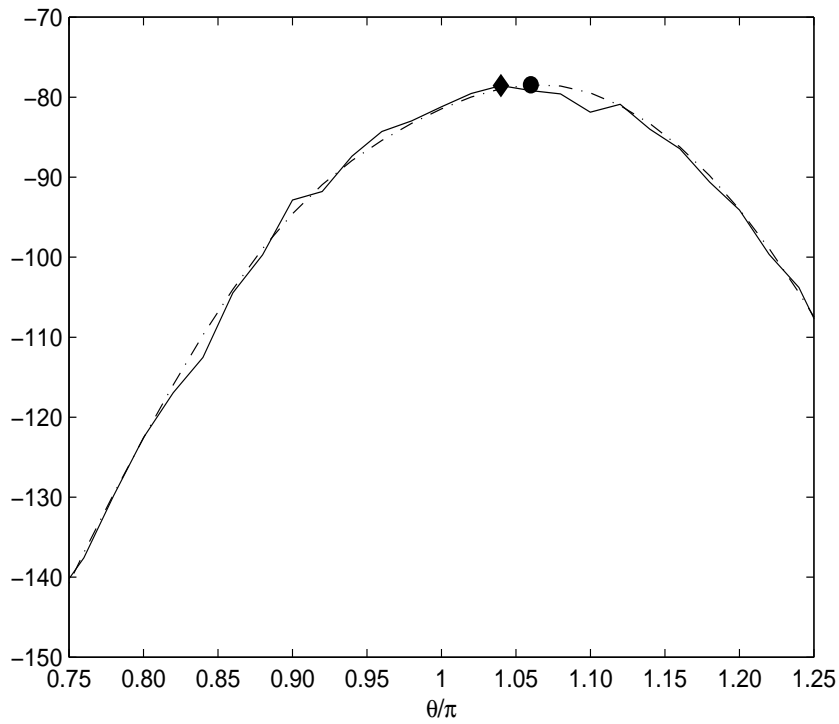
Figure 7: Estimated log-likelihood function using SMC-1 (solid line) and the "true" log-likelihood function (dashed line), with a grid points of every 0.2π, based on a simulated sample path observed at $t = 0, 30, 60, \ldots, 3000$ (101 observations). The corresponding MLEs are shown as the diamond ($\hat{\theta} = 1.04$) for SMC-1 and the circle ($\hat{\theta} = 1.06$) under the 'true' log-likelihood function.

RMSE(θ)	exact sampling	SMC-0	SMC-1
m	80,000	3,500	1,000
$\theta = 0.0\pi$	1.719	0.519	0.325
$\theta = 0.2\pi$	1.488	0.497	0.291
$\theta = 0.4\pi$	1.211	0.433	0.214
$\theta = 0.6\pi$	0.901	0.397	0.157
$\theta = 0.8\pi$	0.648	0.347	0.136
$\theta = 1.0\pi$	0.588	0.331	0.122
$\theta = 1.2\pi$	0.671	0.356	0.135
$\theta = 1.4\pi$	0.870	0.399	0.165
$\theta = 1.6\pi$	1.217	0.452	0.227
$\theta = 1.8\pi$	1.573	0.507	0.299
time(sec.)	0.490	0.478	0.470

Table 4: RMSE for estimating θ using the estimated log-likelihood function under different methods, for the diffusion process with sine drift with 100 simulated sample paths, each observed at $t = 0, 30, 60, \ldots, 3000$ (101 observations) with $\theta = \pi$. Rows 2 reports the Monte Carlo sample sizes m used and the last row reports the average CPU time for estimating the needed log transition density.

3.2 Example 2

Next we consider the jump diffusion process (Merton, 1976; Cox et al., 1979; Aït-Sahalia, 2004)

$$dv_t = (\alpha - \lambda\kappa - \frac{\sigma^2}{2})dt + \sigma dw_t + dz_t, \tag{15}$$

which is often used for modeling stock prices. Here we set $\alpha = 0.08$, $\sigma = 0.2$, $\lambda = 5$ and z_t as a Poisson process with λ being the mean number of arrivals per unit time. In addition, when an Poisson event occurs, z_t produces a jump of size y which follows a normal distribution with mean $\mu_y = 0$ and standard deviation $\sigma_y = 0.1$; $\kappa = E(exp(y) - 1) = 0.005$. A realization of this process is plotted in Figure 8.
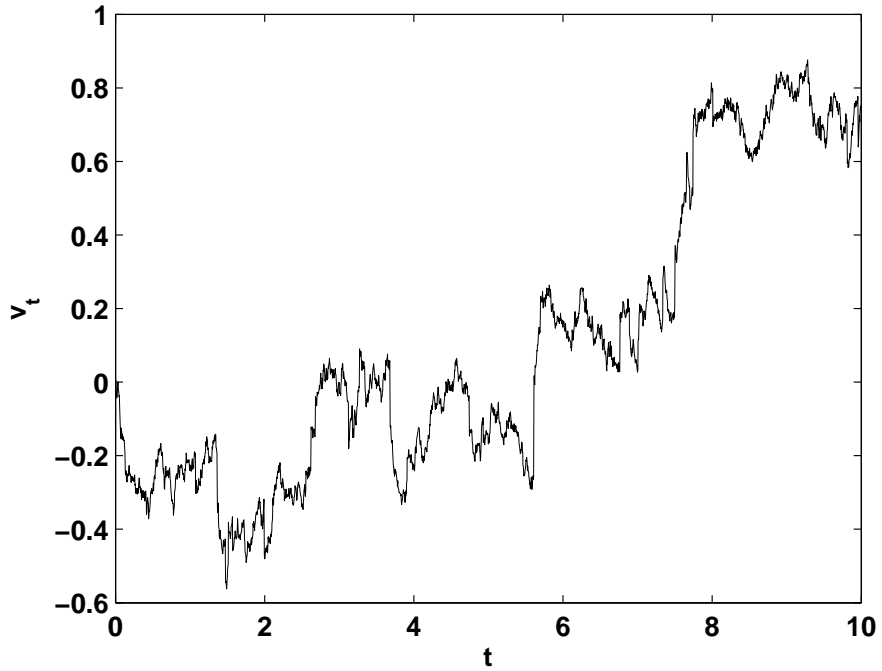
Figure 8: A realization of the jump diffusion process following equation (15) with parameter $\alpha = 0.08$, $\sigma = 0.2$, $\lambda = 5$, $\mu_y = 0$, and $\sigma_y = 0.1$.

For a small time interval δ, let

$$\overline{v}_{t+\delta} = \begin{cases} \overline{v}_t + (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta + w_{t+\delta} - w_t + z_{t+\delta} - z_t & \text{if} \leq 1 \text{ jumps happen in } [t, t+\delta), \\ \overline{v}_t + (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta + w_{t+\delta} - w_t & \text{if} \geq 2 \text{ jumps happen in } [t, t+\delta). \end{cases}$$

28

When we divide $[0, \Delta)$ into M small intervals and let J be the index of the first small interval that more than 1 jumps happen in $[J\delta, (J+1)\delta)$ ($\delta = \Delta/M$), we have $P(J = k) \leq C_1\delta^2$. Hence

$$E(v_\Delta - \overline{v}_\Delta)^2 = \sum_{k=0}^{M-1} P(J = k)E\big[(v_\Delta - \overline{v}_\Delta)^2 \mid J = k\big] \leq C_2\delta.$$

Then the process $\overline{v}_t$ strongly converges to v_t at the rate of $\sqrt{\delta}$. Here C_1 and C_2 are positive constants. Hence, for this process, we can approximate $P(v_k \mid v_{k-1})$ by

$$P^*(v_k \mid v_{k-1}) \sim \begin{cases} \mathcal{N}\big(v_{k-1} + (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta, \ \sigma^2\delta + \sigma_y^2\big) & \text{with probability } 1 - \lambda\delta, \\ \mathcal{N}\big(v_{k-1} + (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta, \ \sigma^2\delta\big) & \text{with probability } \lambda\delta. \end{cases}$$

(16)

3.2.1 Transition density estimation

In this example, the transition probability density $P(V_\Delta \mid V_0)$ only depends on time interval Δ and the difference $V_\Delta - V_0$ between the two end-points. We fixed $V_0 = 0$ and considered the effect of different length of time period $\Delta = i/36$, $i = 1, \cdots, 9$ and different ending point V_Δ. To accommodate the different length of time period, we use different number of intermediate time points M for different Δ. Specifically, we use $M = 100$ for $\Delta = 1/36$, $M = 200$ for $\Delta = 2/36, 3/36, 4/36$, and $M = 400$ for $\Delta = 5/36$ to $9/36$.

To capture the jump behavior of the process, we extend Pedersen (1995)'s sampler as

$$r_k(v_k \mid v_{k-1}) = P^*(v_k \mid v_{k-1}),$$

(17)

and Durham and Gallant (2002)'s sampler as

$$r_k(v_k \mid v_{k-1}) \sim \begin{cases} \mathcal{N}\big(v_{k-1} + \frac{v_M - v_{k-1}}{M-k+1}, \ \frac{M-k}{M-k+1}\sigma^2\delta + \sigma_y^2\big) & \text{with probability } 1 - \lambda\delta, \\ \mathcal{N}\big(v_{k-1} + \frac{v_M - v_{k-1}}{M-k+1}, \ \frac{M-k}{M-k+1}\sigma^2\delta\big) & \text{with probability } \lambda\delta. \end{cases}$$

(18)

It can be verified that both the *extended* Pedersen (1995)'s sampler and the *extended* Durham and Gallant (2002)'s sampler satisfy the consistent condition $\text{Var}_Q(w) < \infty$. For this example, our numerical experiment shows that sampler

29

(17) outperforms sampler (18) under the same computational time. Hence we use (17) as benchmark (denoted by SMC-0). For the proposed procedure SMC-1, we use the same sampling distribution, with the proposed resampling steps. Resampling is performed every two steps, using the resampling priority score (13), with $m^* = 500$ backward pilots. Function f_k is estimated by a simple histogram estimator with partition $[0.04l, 0.04(l+1))$, $l = 0, \pm 1, \pm 2, \cdots$.

The sampling distribution for generating the backward pilots is

$$g_k(u_k \mid u_{k+1}^{(j)}) \propto P^*(u_{k+1}^{(j)} \mid u_k)$$

where P^* is specified in (16). More specifically it is a mixture Gaussian distribution of u_k as follows

$$g_k(u_k \mid u_{k+1}^{(j)}) \sim \begin{cases} \mathcal{N}\big(u_{k+1}^{(j)} - (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta, \ \sigma^2\delta + \sigma_y^2\big) & \text{with probability } 1 - \lambda\delta, \\ \mathcal{N}\big(u_{k+1}^{(j)} - (\alpha - \lambda\kappa - \frac{\sigma^2}{2})\delta, \ \sigma^2\delta\big) & \text{with probability } \lambda\delta, \end{cases}$$

For $\Delta = 1/12$, we plot 100 bridge samples generated by the "perfect" sampling distribution, SMC-0, and SMC-1 in Figure 9 with five consecutive observations $(V_0, V_\Delta, V_{2\Delta}, V_{3\Delta}, V_{4\Delta}) = (0, -0.136, -0.25, -0.27, -0.293)$. It is seen that the sampling distribution under SMC-1 is much closer to the 'perfect' sampling distribution, compared to that under SMC-0. Hence SMC-1 is more efficient.

For transition density estimation, we obtained

$$\text{RMSE}(\Delta, V_\Delta) = \left[\frac{1}{100} \sum_{i=1}^{100} \Big(\log \widehat{P}^{*(i)}(V_\Delta \mid V_0 = 0) - \log P(V_\Delta \mid V_0 = 0) \Big)^2 \right]^{1/2},$$

over 100 independent trials, for various Δ and different V_Δ. In this example, the true value of $\log P(V_\Delta \mid V_0 = 0)$ can be calculated analytically. Estimations obtained by SMC-0 and SMC-1 are based on $m = 5,000$ and $m = 2,000$ Monte Carlo samples, respectively. We plot $\text{RMSE}(\Delta, V_\Delta)$ for $\Delta = 1/36$, $\Delta = 3/36$ and $\Delta = 9/36$ in Figure 10 (Panel (b),(d),(f)), along with the true value of the transition density $P(V_\Delta \mid V_0 = 0)$ (Panel (a),(c),(e)).
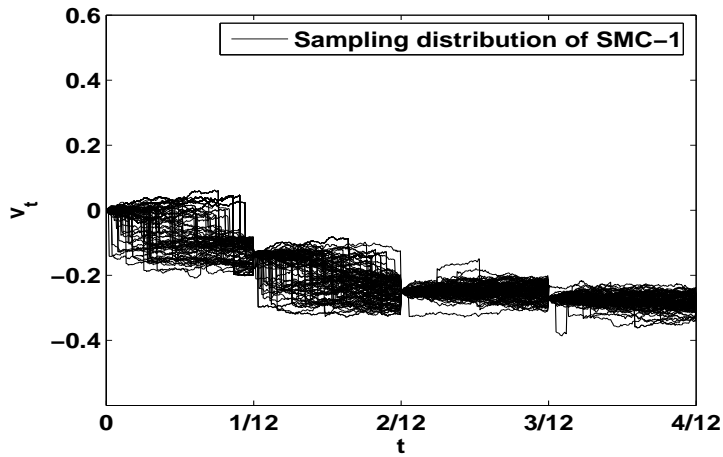
Table 5 reports the overall performance measure

$$\text{RMSE}(\Delta) \triangleq \left[\int \text{RMSE}^2(\Delta, V_\Delta) \, P(V_\Delta \mid V_0 = 0) \, dV_\Delta \right]^{1/2},$$
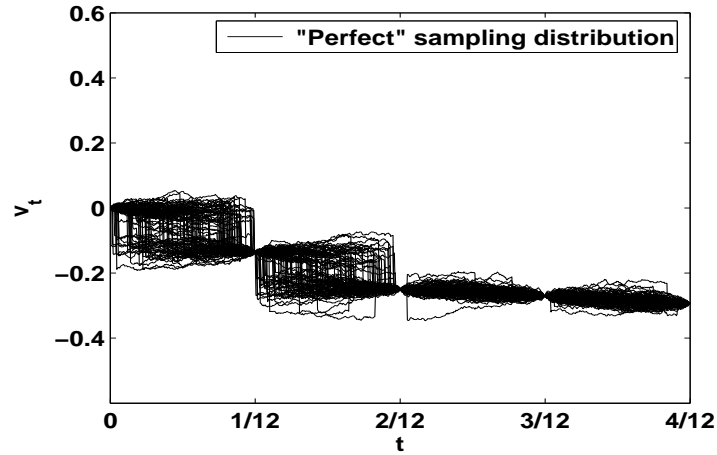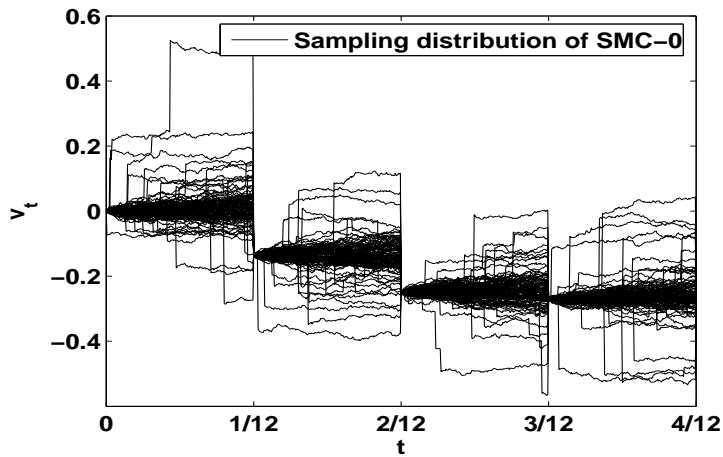
Figure 9: Bridge samples generated using different methods for the jump diffusion process with observe time interval $\Delta = 1/12$. The observations are $V_0 = 0$, $V_{1/12} = -0.136$, $V_{2/12} = -0.250$, $V_{3/12} = -0.270$ and $V_{4/12} = -0.293$. (a): The "perfect" sampling distribution; (b): Extended Pedersen (1995)'s sampling method without resampling (SMC-0); (c): Extended Pedersen (1995)'s sampling method with resampling steps (SMC-1). In SMC-1, the resampling step is performed every two steps when generating the bridge samples. $m^* = 500$ backward pilots are generated to estimate the resampling priority scores.
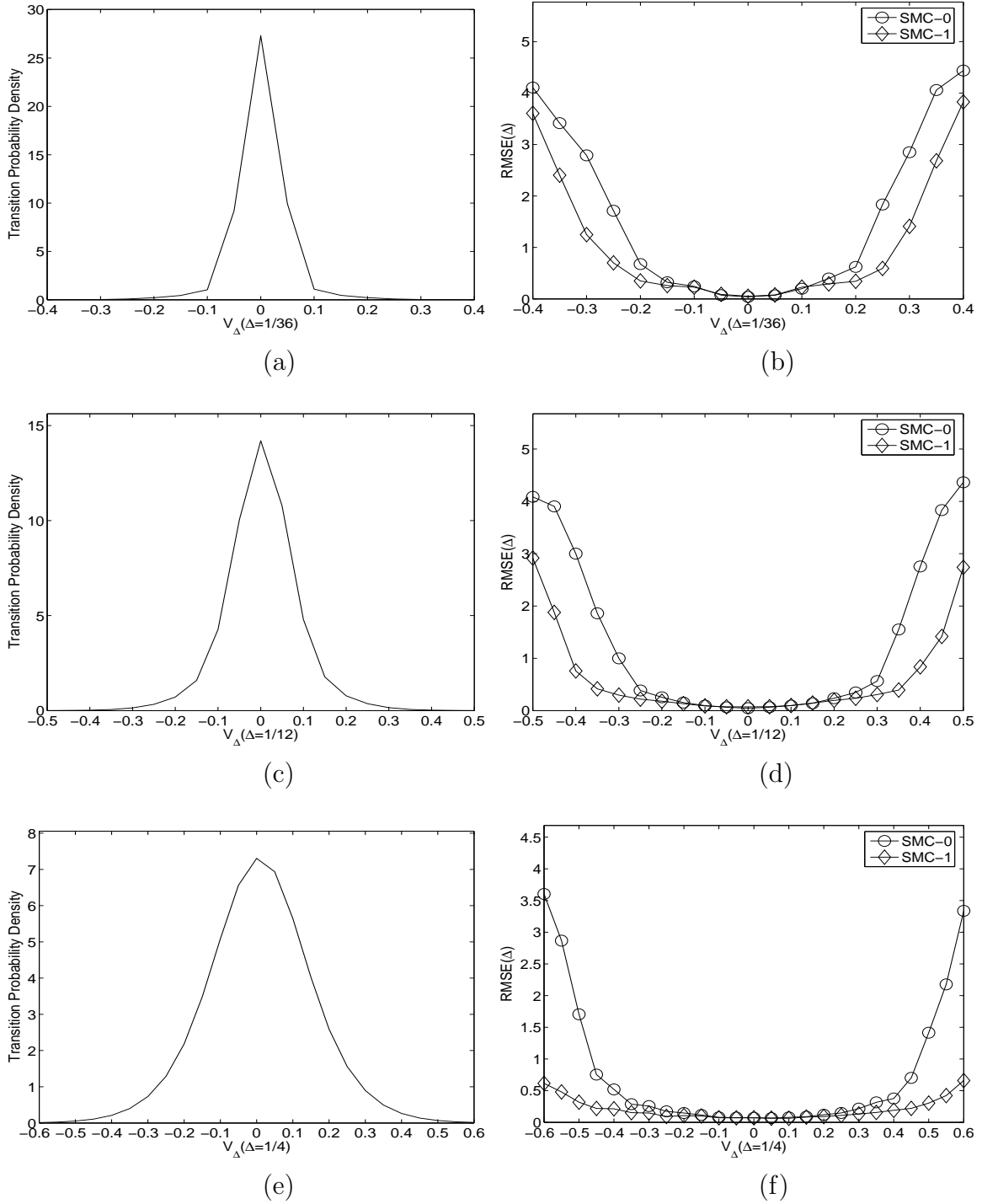
Figure 10: True transition density $P(V_\Delta \mid V_0 = 0)$ (Panels (a),(c),(e)) for the jump diffusion process with observation time intervals $\Delta = 1/36$, $\Delta = 1/12$ and $\Delta = 1/4$, respectively. Panels (b),(d) and (f) show the corresponding $\text{RMSE}(\Delta, V_\Delta)$ of estimating the transition densities using SMC-0 with $m = 5,000$ samples and SMC-1 with $m = 2,000$ samples.

32

RMSE(Δ) of log-likelihood	SMC-0	SMC-1
m	5,000	2,000
$\Delta = 1/36$	0.187	0.129
$\Delta = 2/36$	0.211	0.125
$\Delta = 3/36$	0.193	0.113
$\Delta = 4/36$	0.190	0.105
$\Delta = 5/36$	0.236	0.119
$\Delta = 6/36$	0.226	0.114
$\Delta = 7/36$	0.218	0.111
$\Delta = 8/36$	0.215	0.109
$\Delta = 9/36$	0.211	0.108
time(sec.)	0.209	0.203

Table 5: RMSE(Δ) of using SMC-0 and SMC-1 to estimate the log-transition probability of the jump diffusion process. Row 2 is the Monte Carlo sample sizes (m) and the last row reports the average CPU time of each evaluation of the log transition density.

while controlling the CPU time to be roughly the same for different methods. It is clear that SMC-1 is more efficient.

3.2.2 Estimation of the realized volatility

Another interesting use of diffusion bridge samples is to estimate the realized volatility (Hull and White, 1987; Barndorff-Nielsen and Shephard, 2002; Zhang et al., 2005) conditional on the two end-points. In this example, when properly weighted samples of $(v_0^{(j)} = V_0, v_1^{(j)}, \ldots v_{M-1}^{(j)}, v_M = V_\Delta)$ are obtained, we can estimate $I(\Delta, V_\Delta) = E(\sum_{s=1}^{M} (v_s - v_{s-1})^2 \mid V_0, V_\Delta)$ by

$$\widehat{I}(\Delta, V_\Delta) = \frac{\sum_{j=1}^{m} w^{(j)} \sum_{s=1}^{M} (v_s^{(j)} - v_{s-1}^{(j)})^2}{\sum_{j=1}^{m} w^{(j)}}. \tag{19}$$

In fact, there is a better resampling the priority score for this specific inference

problem. The "optimal" resampling priority score developed in (11) was only designed to generate the "best" bridge samples and to minimize variance of weight $\text{Var}_Q w$.

At step k ($k \leq M-2$), when constructing the priority score $\beta_k^{(j)}$ for estimating $I(\Delta, V_\Delta)$, we consider minimizing the variance of $w^{(j)}(\sum_{s=1}^{M}(v_s^{(j)} - v_{s-1}^{(j)})^2)^{1/2}$ instead of minimizing the variance of $w^{(j)}\sum_{s=1}^{M}(v_s^{(j)} - v_{s-1}^{(j)})^2$, which is the term $\text{Var}_Q(w\,h)$ (or $\text{Var}_Q(\overline{w}\,h)$ equivalently) in the MSE approximation (7). Again, given the forward sampling distributions $r_s(v_s \mid v_{s-1}, v_M; \theta)$, $s = k+1, \cdots, M-1$, without considering the effect of future resampling after step k, the problem is equivalent to finding β that minimizes

$$
E\left[w^2 \sum_{s=1}^{M}(v_s - v_{s-1})^2 \,\Big|\, V_0, V_\Delta\right]
$$

$$
= E\left[\left(\frac{\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)}{\beta_k Q_k(\boldsymbol{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)}\right)^2 \sum_{s=1}^{M}(v_s - v_{s-1})^2 \,\Big|\, V_0, V_\Delta\right]
$$

$$
= \int \frac{\left[\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\beta_k Q_k(\boldsymbol{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \sum_{s=1}^{M}(v_s - v_{s-1})^2 \, dv_1 \cdots dv_{M-1}
$$

$$
= \int \frac{\left[\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\beta_k Q_k(\boldsymbol{v}_k)} \sum_{s=1}^{k}(v_s - v_{s-1})^2 \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} dv_{k+1} \cdots dv_{M-1} dv_1 \cdots dv_k
$$

$$
+ \int \frac{\left[\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\beta_k Q_k(\boldsymbol{v}_k)} \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \sum_{s=k+1}^{M}(v_s - v_{s-1})^2 \, dv_{k+1} \cdots dv_{M-1} dv_1 \cdots dv_k.
$$

Hence, $\text{Var}\left[w\sqrt{\sum_{s=1}^{M}(v_s - v_{s-1})^2}\right]$ is minimized when

$$
\beta_k = w_k \left[\sum_{s=1}^{k}(v_s - v_{s-1})^2 \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} dv_{k+1} \cdots dv_{M-1}\right.
$$

$$
\left. + \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \sum_{s=k+1}^{M}(v_s - v_{s-1})^2 \, dv_{k+1} \cdots dv_{M-1}\right]^{1/2}.
$$

The backward pilot scheme described in Section 2.3 can be used to estimate

$$
\widehat{f}_k^{(1)}(v_k; \theta) \simeq \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} dv_{k+1} \cdots dv_{M-1},
$$

and

$$\widehat{f}_k^{(2)}(v_k;\theta) \simeq \int \frac{\left[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1};\theta)\right]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M;\theta)} \sum_{s=k+1}^{M} (v_s - v_{s-1}) \, dv_{k+1} \cdots dv_{M-1}.$$

Then the priority score is set as

$$\beta_k \propto w_k \left[\sum_{s=1}^{k} (v_s - v_{s-1})^2 \widehat{f}_k^{(1)}(v_k;\theta) + \widehat{f}_k^{(2)}(v_k;\theta) \right]^{1/2}. \tag{20}$$

We denote SMC using the resampling priority score (20) as SMC-1*, while SMC-1 uses (11) as the resampling priority score. The other settings in SMC-1* are the same as in SMC-1. For performance comparison, define

$$\mathrm{RMSE}_\sigma(\Delta, V_\Delta) \;=\; \left[\frac{1}{100} \sum_{i=1}^{100} \left(\widehat{I}^{(i)}(\Delta, V_\Delta) - I(\Delta, V_\Delta) \right)^2 \right]^{1/2},$$

over 100 independent estimates as the measurement of estimate accuracy, where the "true" value of $I(\Delta, V_\Delta) = E(\sum_{s=1}^{M} (v_s - v_{s-1})^2 \mid V_0, V_\Delta)$ is obtained by using SMC-0 with $m = 500,000$ Monte Carlo samples. $\mathrm{RMSE}_\sigma(\Delta, V_\Delta)$ for $\Delta = 1/36$, $\Delta = 3/36$ and $\Delta = 9/36$ are plotted in Figure 11. Figure 11 also shows the "true" value of $E(\sum_{s=1}^{M} (v_s - v_{s-1})^2 \mid V_0, V_\Delta)$. It can be seen that the SMC methods with backward pilot guided resampling perform well, especially for longer time period Δ and when the difference between the two end-points $|V_\Delta - V_0|$ is large.

Table 6 reports the overall performance measurement defined as

$$\mathrm{RMSE}_\sigma(\Delta) = \left[\int \mathrm{RMSE}_\sigma^2(\Delta, V_\Delta) \, P(V_\Delta \mid V_0 = 0) \, dV_\Delta \right]^{1/2},$$

under similar CPU time. It appears that the specifically designed resampling priority scores indeed improve the performance over the generic resampling priority scores.

4 Conclusions and Discussion

In this paper, we proposed a resampling scheme under SMC framework to generate samples of diffusion bridge that connect the two observed ends of the underlying diffusion process. This resampling scheme can be easily combined with many other
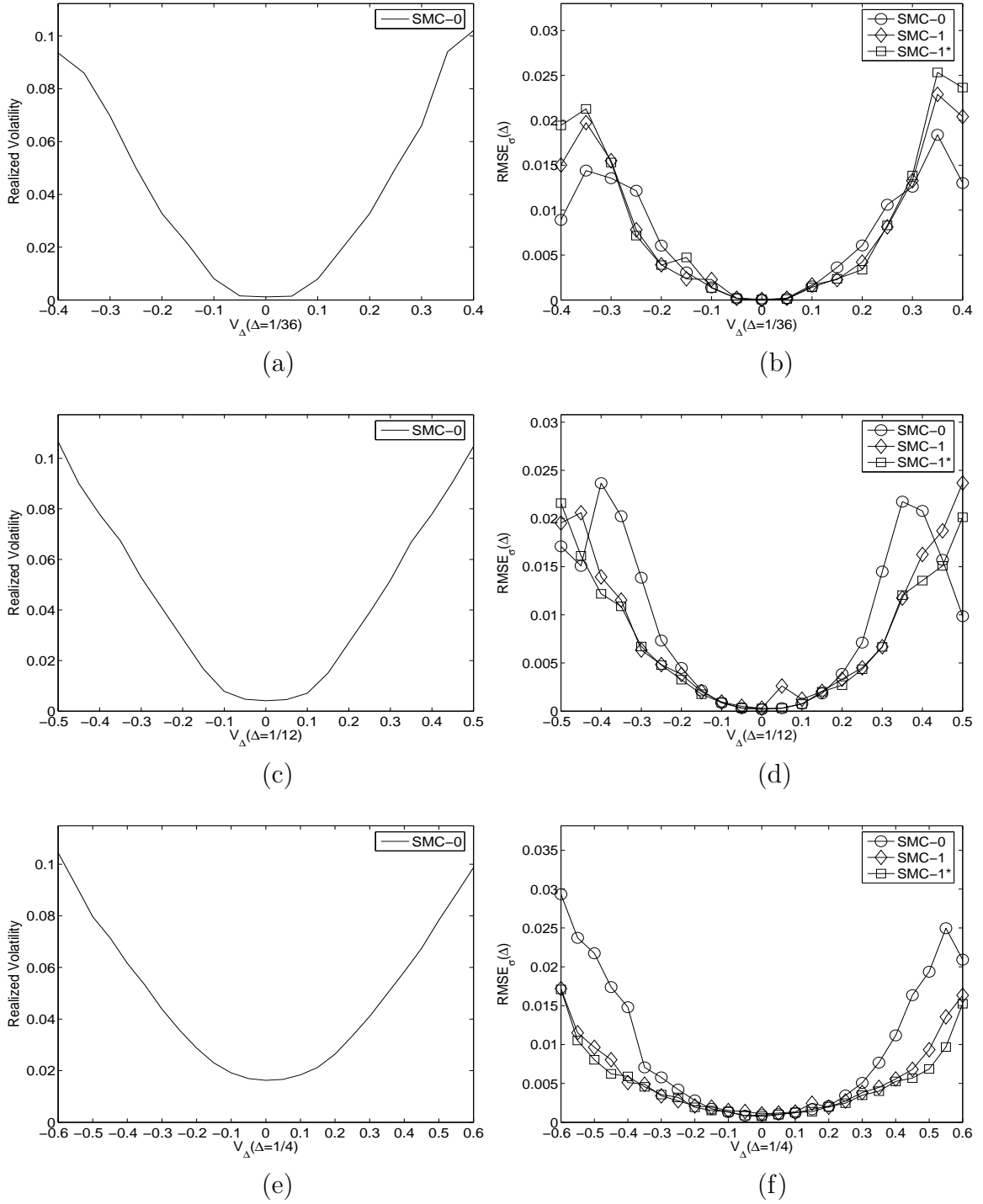
Figure 11: Plots of the "true" values of $E(\sum_{s=1}^{M}(v_s - v_{s-1})^2 \mid V_0 = 0, V_\Delta)$ (Panel (a),(c),(e)) of the jump diffusion process with observation time intervals $\Delta = 1/36$, $\Delta = 1/12$ and $\Delta = 1/4$, respectively. Panels (b),(d) and (f) show the corresponding $\text{RMSE}_\sigma(\Delta, V_\Delta)$ using SMC-0 with $m = 5,000$ samples, SMC-1 with $m = 2,000$ samples, and SMC-1* with $m = 2,000$ samples. In SMC-1 and SMC-1* , the resampling step is performed every two steps when generating the bridge samples. $m^* = 500$ backward pilots are generated to estimate the resampling priority scores. 36

RMSE$_\sigma(\Delta)$ ($\times 10^2$)	SMC-0	SMC-1	SMC-1*
m	5,000	2,000	2,000
$\Delta = 1/36$	0.119	0.106	0.105
$\Delta = 2/36$	0.177	0.135	0.132
$\Delta = 3/36$	0.200	0.165	0.140
$\Delta = 4/36$	0.211	0.163	0.152
$\Delta = 5/36$	0.262	0.186	0.168
$\Delta = 6/36$	0.272	0.196	0.177
$\Delta = 7/36$	0.283	0.203	0.190
$\Delta = 8/36$	0.296	0.212	0.191
$\Delta = 9/36$	0.310	0.225	0.200
time(sec.)	0.209	0.203	0.211

Table 6: RMSE$_\sigma(\Delta)$ for using different sampling methods to estimate $E(\sum_{s=1}^{M}(v_s - v_{s-1})^2 \mid V_0, V_\Delta)$ of the jump diffusion process. Row 2 is the Monte Carlo sample sizes (m) and the last row reports the average CPU time of each evaluation.

sampling methods, including Pedersen (1995)'s sampling method and Durham and Gallant (2002)'s sampling method. The resampling steps adjust the sampling distribution in the intermediate steps according to priority scores. The priority score of the resampling is optimized to minimize the Chi-square divergence between the target distribution and the sampling distribution. Since the analytic values of the "optimal" priority scores cannot be computed easily, backward pilots are generated to estimate the values. The computational complexity of computing the "optimal" priority scores is limited. Two synthetic examples are used to demonstrate the effectiveness of this resampling scheme.

For different purposes of generating diffusion bridge samples, the "optimal" priority scores can be different. In cases of parameter estimations, likelihood function estimates are often expected to be continuous (Pitt, 2002). More research is needed to apply this resampling scheme to estimating parameters.

References

Aït-Sahalia, Y. (2004), "Disentangling diffusion from jumps," *Journal of Financial Economics*, 74, 487–528.

Barndorff-Nielsen, O. E. and Shephard, N. (2002), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society, Series B*, 64, 253C280.

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006), "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes," *Journal of Royal Statistical Society B*, 68, 333–382.

Beskos, A., Roberts, G., Stuart, A., and Voss, J. (2008), "MCMC methods for diffusion bridges," Tech. rep., University of Warwick.

Brandt, M. W. and Santa-Clara, P. (2002), "Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets," *Journal of Financial Economics*, 63, 61–210.

Cox, J., Ross, S., and Rubinstein, M. (1979), "Option pricing : A simplified approach," *Journal of Financial Economics*, 7, 229–263.

Durham, G. B. and Gallant, A. R. (2002), "Numerical Techniques for Maximum Likelihood Estimation of Continuous-time Diffusion Processes," *Journal of Business & Economic Statistics*, 20, 297–338.

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," *Journal of Business & Economic Statistics*, 19, 177–191.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1995), *Markov chain Monte Carlo in Practice*, Chapman & Hall.

Hull, J. and White, A. (1987), "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, 42, 281C300.

Kloeden, P.E. and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin Heidelberg New York.

Kong, A., Liu, J., and Wong, W. (1994), "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, 89, 278–288.

Liu, J. and Chen, R. (1998), "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, 93, 1032–1044.

Liu, J. S. (1996), "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Compustion*, 6, 113–119.

— (2001), *Monte Carlo Strategies in Scientific Computing*, Springer, New York.

Marshall, A. (1956), "The use of multi-stage sampling schemes in Monte Carlo computations," in *Symposium on Monte Carlo Methods*, ed. Meyer, M., Wiley, pp. 123–140.

Merton, R. C. (1976), "Option pricing when underlyinh stock returns are discontinous," *Journal of Financial Economics*, 3, 125–144.

Pedersen, A. R. (1995), "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 257–279.

Pitt, M. K. (2002), "Smooth particle filters for likelihood and Maximisation," Tech. rep., University of Warwick.

Robert, R. and Casella, G. (1999), *Monte Carlo statistical methods*, Springer, New York.

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

Shoji, I. and Ozaki, T. (1998), "Estimation for Nonlinear Stochastic Differential Equations by a Local Linearization Method," *Stochastic Analysis and Applications*, 16, 733–752.

Stramer, O. and Yan, J. (2006), "On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation," Tech. rep., University of Iowa.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005), "A tale of two time scales: determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.