On Generating Monte Carlo Samples of Continuous Diffusion Bridges

Ming LIN, Rong CHEN, and Per MYKLAND

Diffusion processes are widely used in engineering, finance, physics, and other fields. Usually continuous-time diffusion processes can be observed only 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 for generating the intermediate paths of the bridge. The paths often are generated forward from the starting observation and forced in some ways to connect with the end observation. In this article we propose a constrained SMC algorithm with an effective resampling scheme 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.

KEY WORDS: Backward pilot; Priority score; Resampling; Sequential Monte Carlo; Stochastic diffusion equation.

1. INTRODUCTION

Diffusion processes are widely used in engineering, finance, physics, and many other fields. In practice, a diffusion process often is observable only at discrete time points. On the other hand, for most nonlinear and non-Gaussian diffusion processes, statistical inferences can be carried out much more easily 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. Because of the Markovian nature of diffusion processes, the imputation problem becomes one of generating continuous paths of the underlying diffusion process that connect two fixed endpoints (diffusion bridges). In this article 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),

$$dV_t = \mathbf{b}(V_t;\boldsymbol{\theta})\,dt + \mathbf{A}(V_t;\boldsymbol{\theta})\,d\mathbf{W}_t, \tag{1}$$

where $\mathbf{W}_t = (w_{t,1},\ldots,w_{t,d})^T$ are d independent Brownian motions, $\mathbf{b}(V_t;\boldsymbol{\theta}) = (b_1(V_t;\boldsymbol{\theta}),\ldots,b_d(V_t;\boldsymbol{\theta}))^T$ are the drift coefficients, $\mathbf{A}(V_t;\boldsymbol{\theta}) = \{a_{i,j}(V_t;\boldsymbol{\theta})\}_{d\times d}$ are the diffusion coefficients, and $\boldsymbol{\theta}$ is the parameter in the coefficients. For notational simplicity, herein we use V_t instead of $\mathbf{V}_t$, although V_t is a d-dimensional vector.

The methods that we develop in this work can be easily extended to time-inhomogeneous processes, in which the drift coefficients and diffusion coefficients also may depend on the time variable t. In addition, our methods also apply to jump diffusion processes such as

$$dV_t = \mathbf{b}(V_t;\boldsymbol{\theta})\,dt + \mathbf{A}(V_t;\boldsymbol{\theta})\,d\mathbf{W}_t + d\mathbf{Z}_t,$$

where $\mathbf{Z}_t$ 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, here we concentrate on the diffusion process (1).

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

Beskos et al. (2006) proposed generating continuous sample paths exactly following the conditional distribution $P(V_t \mid V_0, V_\Delta;\boldsymbol{\theta})$ 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 probability density $P(V_\Delta \mid V_0,\boldsymbol{\theta})$. This method was shown to work well for many processes, although it is limited to reducible processes, and sometimes the acceptance rate can be very small for certain situations, especially when Δ is large.

Several simulation methods based on the discrete-time approximation of diffusion process have been developed (Kloeden and Platen 1992; Pedersen 1995; Elerian, Chib, and Shephard 2001; Eraker 2001; Roberts and Stramer 2001; Brandt and Santa-Clara 2002; Durham and Gallant 2002). 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 $(V_{t_0} = V_0, V_{t_1},\ldots,V_{t_{M-1}}, V_{t_M} = V_\Delta)$ following the distribution

$$P^*\big(V_{t_1},\ldots,V_{t_{M-1}} \,\big|\, V_{t_0} = V_0, V_{t_M} = V_\Delta;\boldsymbol{\theta}\big) \propto \prod_{k=1}^{M} P_\delta^*\big(V_{t_k} \,\big|\, V_{t_{k-1}};\boldsymbol{\theta}\big), \tag{2}$$

where

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

Ming Lin is Associate Professor, the Wang Yanan Institute for Studies in Economics, Xiamen University. Rong Chen is Professor of Statistics, Department of Statistics, Rutgers University, Piscataway, NJ 08854 and Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University, Beijing 100871, China (E-mail: *rongchen@stat.rutgers.edu*). Per Mykland is Professor of Statistics, University of Chicago, Chicago, IL 60636. Chen's research is sponsored in part by NSF grant DMS 0800183 and Oxford–Man Institute of Quantitative Finance, Oxford University. Mykland's research is supported in part by NSF grants DMS 06-04758 and SES 06-31605. The authors thank an associate editor and three anonymous referees for their helpful comments.

Journal of the American Statistical Association
June 2010, Vol. 105, No. 490, Theory and Methods
DOI: 10.1198/jasa.2010.tm09057

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

Generating Monte Carlo samples from the target distribution (2) can be done using Markov chain Monte Carlo (MCMC) through a transition kernel whose equilibrium distribution is the target distribution (Gilks, Richardson, and Spiegelhalter 1995; Robert and Casella 1999). However, for most diffusion processes and for large M (to achieve approximation accuracy), the mixing rate of MCMC can be very low. To avoid this problem, Roberts and Stramer (2001) and Elerian, Chib, and Shephard (2001) proposed updating a block of the bridge sample in one MCMC move, with the proposed 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 $P^*(V_{t_M} \mid V_{t_0}; \boldsymbol{\theta})$ for the target distribution (2).

In the present 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 $V_{t_1}^{(j)}, V_{t_2}^{(j)}, \ldots, V_{t_{M-1}}^{(j)}$ are generated sequentially until the complete bridge samples $(V_{t_0}, V_{t_1}^{(j)}, \ldots, V_{t_{M-1}}^{(j)}, V_{t_M})$ are obtained. The critical issue here is how to utilize the information provided by the end observation V_{t_M} when generating the intermediate states $(V_{t_1}, \ldots, V_{t_{M-1}})$.

An SMC approach for generating diffusion bridges was proposed and studied by (Pedersen 1995; Brandt and Santa-Clara 2002; Durham and Gallant 2002). Their approaches have been shown to be effective in some cases, but to 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 article we propose an effective resampling scheme in SMC. The resampling is guided by pilots 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 the samplers of Pedersen (1995) and Durham and Gallant (2002).

The rest of the article is organized as follows. Section 2 introduces the proposed algorithm, with brief backgrounds on SMC, the 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. Section 4 concludes.

2. CONSTRAINED SEQUENTIAL MONTE CARLO GUIDED BY BACKWARD RESAMPLING

In what follows, 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. Again, we use v_k instead of $\mathbf{v}_k$ even though v_k is a d-dimentional vector.

2.1 Importance Sampling and Sequential Monte Carlo

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

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

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

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

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

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

is a consistent estimator of the expectation of $h(\mathbf{v})$ conditional on the given end-points 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 $Q(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$. When the sampling distribution is "perfect," that is,

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

the estimator (4) provides the exact value of $P^*(v_M \mid v_0; \boldsymbol{\theta})$. Although directly generating samples from $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ and calculating the weights is not feasible in most cases, an efficient sampling distribution $Q(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ should be close to $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$. For the purpose of efficiency control, the chi-squared divergence between $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ and $Q(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$, defined as

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

is often used as a performance measure of the chosen sampling distribution $Q(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ (Liu 2001). Here $\overline{w} = w/P^*(v_M \mid$

$v_0; \boldsymbol{\theta})$ is the *standardized weight*. Assuming that the samples generated are independent, the mean squared error (MSE) of estimator (5) can be approximated by (Kong, Liu, and Wong 1994; Liu 1996)

$$\frac{1}{m}\big[E^2_{P^*}(h \mid v_0, v_M; \boldsymbol{\theta})\,\mathrm{Var}_Q(\overline{w}) + \mathrm{Var}_Q(\overline{wh}) - 2E_{P^*}(h \mid v_0, v_M; \boldsymbol{\theta})\,\mathrm{Cov}_Q(\overline{w}, \overline{wh})\big]. \quad (7)$$

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

Elerian, Chib, and Shephard (2001) proposed using the multivariate normal distribution or the multivariate Student t-distribution as $Q(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$. When M is large, directly constructing a good sampling distribution close to the target distribution $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ in such a high-dimensional space is usually difficult.

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

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

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

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

1. Let $\mathbf{v}_0^{(j)} = \{v_0, v_M\}$ and $w_0^{(j)} = 1$.
2. For $k = 1, \ldots, M-1$,
   (a) Draw $v_k^{(j)}$ from distribution $r_k(v_k \mid \mathbf{v}_{k-1}^{(j)}; \boldsymbol{\theta})$. Let $\mathbf{v}_k^{(j)} = \{\mathbf{v}_{k-1}^{(j)}, v_k^{(j)}\}$.
   (b) Compute the corresponding weight of $\mathbf{v}_k^{(j)}$ by
$$w_k^{(j)} = w_{k-1}^{(j)} \frac{P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \boldsymbol{\theta})}{r_k(v_k^{(j)} \mid \mathbf{v}_{k-1}^{(j)}; \boldsymbol{\theta})}.$$
3. Let $\mathbf{v}_M^{(j)} = \mathbf{v}_{M-1}^{(j)}$ and $w^{(j)} = w_{M-1}^{(j)} P^*(v_M \mid v_{M-1}^{(j)}; \boldsymbol{\theta})$.

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

Pedersen (1995) proposed using the forward equation

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

so that $w_k^{(j)} \equiv 1$ for $k = 1, \ldots, M-1$, and $w^{(j)} = P^*(v_M \mid v_{M-1}^{(j)}; \boldsymbol{\theta})$. The procedure essentially involves 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 endpoint 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 "jumps" between $v_{M-1}^{(j)}$ and the fixed endpoint v_M. In many cases, the performance of this simple sampling method is not satisfactory.

Durham and Gallant (2002) suggested a different sampling distribution,

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

This proposal distribution includes a drift term that linearly connects the current position v_{k-1} to the targeted endpoint v_M, thereby forcing v_k to move toward v_M as k increases to M. Stramer and Yan (2006) proved that this sampling distribution is the "perfect" sampling distribution $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ when both the drift coefficient $\mathbf{b}(v_t; \boldsymbol{\theta})$ and the diffusion coefficient $\mathbf{A}(v_t; \boldsymbol{\theta})$ do not depend on v_t. But this sampling distribution might not be ideal for some applications, especially when the drift coefficient depends strongly on v_t or when the time interval Δ is large.

Consider, for example, the Ornstein–Uhlenbeck process

$$dv_t = \theta v_t\,dt + dw_t. \quad (10)$$

Let $\theta = 0.2$. Figure 1 shows 100 sample paths (without taking into account the weight) generated from the true conditional distribution $P^*(\mathbf{v} \mid v_0 = 0, v_M = 28.3)$ and from Durham and Gallant (2002)'s sampling distribution (9), with time interval $\Delta = 20$ and $M = 400$ intermediate points. Clearly, 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, Chib, and Shephard (2001) documented similar findings.

2.2 Resampling and Optimal Resampling Priority Score

Resampling (Kong, Liu, and Wong 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 buildup of the samples. It provides a way to rejuvenate the samples to improve efficiency in future steps. In our problem the target distribution $P^*(\mathbf{v} \mid v_0, v_M; \boldsymbol{\theta})$ dictates that the samples of $\mathbf{v}$ must connect two fixed points, v_0 and v_M, thus imposing a very strong constraint on the sample path. During the sequential buildup, if a partial sample path $\mathbf{v}_k$ 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 buildup to complete the sample, because the complete sample would have very small weight and make a negligible contribution to the weighted average used for statistical inferences.

Resampling is done as follows. Suppose that we have obtained samples $\{(\mathbf{v}_k^{(j)}, w_k^{(j)}), j = 1, \ldots, m\}$ at step k. The resampling step creates a new set of samples, $\{(\mathbf{v}_k^{\mathrm{new}(j_*)}, w_k^{\mathrm{new}(j_*)}), j_* = 1, \ldots, m\}$, by drawing samples from the current set $\{\mathbf{v}_k^{(j)}, j = 1, \ldots, m\}$ with replacement according to *priority*

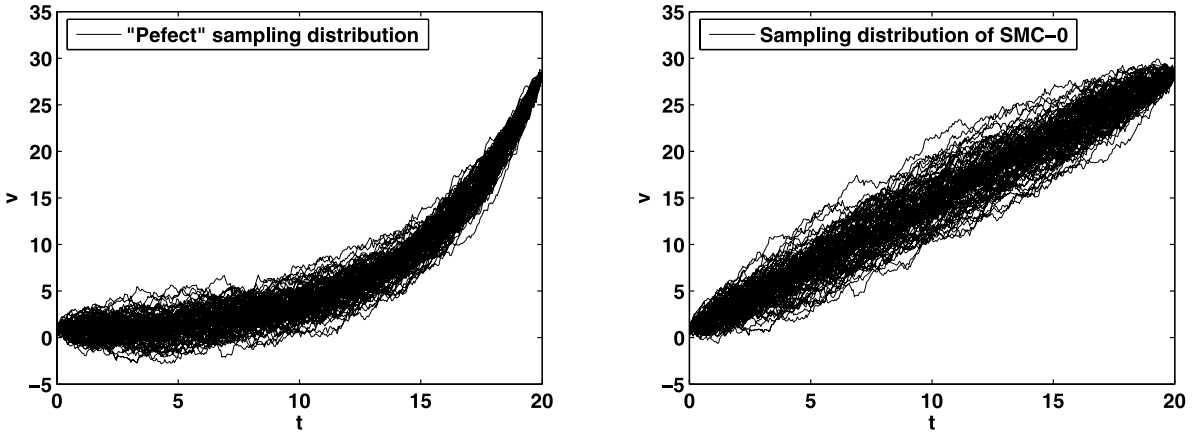


Figure 1. Samples of diffusion bridges generated from (a) the true conditional distribution and (b) Durham and Gallant (2002)'s sampling method for diffusion process (10) with parameter $\theta = 0.2$. $M = 400$ intermediate points are used.

scores $\{\beta_k^{(j)}, j = 1, \ldots, m\}$, and adjusting the weights accordingly so that (Liu and Chen 1998)

$$E\left[\frac{1}{m}\sum_{j_*=1}^{m} w_k^{\text{new}(j_*)} h\big(\mathbf{v}_k^{\text{new}(j_*)}\big) \,\Big|\, \mathbf{v}_k^{(j)}, w_k^{(j)}, j = 1, \ldots, m\right] = \frac{1}{m}\sum_{j=1}^{m} w_k^{(j)} h\big(\mathbf{v}_k^{(j)}\big)$$

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

1. Assign a priority score $\beta_k^{(j)} > 0$ to each sample $\mathbf{v}_k^{(j)}$. Normalize the priority scores so that $\sum_{j=1}^m \beta_k^{(j)} = m$.
2. For $j_* = 1, \ldots, m$,
   (a) Randomly sample $\mathbf{v}_k^{\text{new}(j_*)}$ from the set $\{\mathbf{v}_k^{(j)}, j = 1, \ldots, m\}$ with probabilities proportional to the priority scores $\{\beta_k^{(j)}, j = 1, \ldots, m\}$;
   (b) If $\mathbf{v}_k^{\text{new}(j_*)} = \mathbf{v}_k^{(j)}$, then set the new weight associated with $\mathbf{v}_k^{\text{new}(j_*)}$ to be $w_k^{\text{new}(j_*)} = w_k^{(j)}/\beta_k^{(j)}$.
3. Return the new set of weighted samples $\{(\mathbf{v}_k^{\text{new}(j_*)}, w_k^{\text{new}(j_*)}), j_* = 1, \ldots, m)\}$.

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 a sample's "goodness." In what follows, we develop the "optimal" priority score for our diffusion bridge problem. Note that the resampling step is designed to improve the efficiency of future steps, resampling at step $M-1$ is not needed according to Rao–Blackwellization (Liu and Chen 1998). Thus we only develop the "optimal" priority score, β_k, for step $k = 1, \ldots, M-2$. Again, our goal is to minimize $\text{Var}_Q(w)$ in (6).

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

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

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

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

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

$$\begin{aligned}
&\text{Var}_Q(w)\\
&= E\left[\frac{\prod_{s=1}^M P^*(v_s \mid v_{s-1}; \boldsymbol{\theta})}{\beta_k Q_k(\mathbf{v}_k)\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \boldsymbol{\theta})}\right]^2 - (P^*(v_M \mid v_0; \boldsymbol{\theta}))^2\\
&= \int \frac{[\prod_{s=1}^M P^*(v_s \mid v_{s-1}; \boldsymbol{\theta})]^2}{\beta_k Q_k(\mathbf{v}_k)\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \boldsymbol{\theta})}\, dv_1 \cdots dv_{M-1} - (P^*(v_M \mid v_0; \boldsymbol{\theta}))^2\\
&= \int \frac{[\prod_{s=1}^k P^*(v_s \mid v_{s-1}; \boldsymbol{\theta})]^2}{\beta_k Q_k(\mathbf{v}_k)} \times \int \frac{[\prod_{s=k+1}^M P^*(v_s \mid v_{s-1}; \boldsymbol{\theta})]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \boldsymbol{\theta})}\, dv_{k+1} \cdots dv_{M-1}\, dv_1 \cdots dv_k\\
&\quad - (P^*(v_M \mid v_0; \boldsymbol{\theta}))^2.
\end{aligned}$$

Thus $\text{Var}_Q(w)$ is minimized when

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

That is,

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

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

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

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

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

In this case, the priority score is proportional to the transition probability from the current position v_k to the fixed endpoint v_M, although evaluating this quantity faces the same difficulty as in 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

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

One way to draw samples 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 faced when obtaining the optimal priority score assignment (11) is that the value of integration,

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

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

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

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

1. For $k = M$, let $a_M^{(j)} = 1$, $u_M^{(j)} = v_M$, $j = 1, \ldots, m^*$.
2. For $k = M-1, M-2, \ldots, 1$, for each $j = 1, \ldots, m^*$:
   (a) Generate $u_k^{(j)}$ from a sampling distribution $g_k(u_k \mid u_{k+1}^{(j)}; \boldsymbol{\theta})$. The choice of $g_k(u_k \mid u_{k+1}^{(j)}; \boldsymbol{\theta})$ is discussed in Remark 2.
   (b) Calculate the corresponding weight of $\mathbf{u}_k^{(j)}$ by

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

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

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

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

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

   where

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

   $|\mathbb{D}_{k,l}|$ is the volume of subspace $\mathbb{D}_{k,l}$, and $\mathbb{I}(\cdot)$ is the indicator function.
   (d) (Optional) Perform resampling to the backward pilots $\{(\mathbf{u}_k^{(j)}, a_k^{(j)}), j = 1, \ldots, m^*\}$, 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) were generated up to time $k = 100$. The backward pilots samples up to $k = 100$ are shown, as is the histogram estimate of f_k at $k = 100$ based on these backward pilot samples (vertically) on the right side of the figure. It can be seen that path (b), as a partial sample of the bridge, is a better sample, because it has higher probability of connecting to the fixed end compared with path (a).

Remarks.

1. The backward pilots $\mathbf{u}_k^{(j)}$, $j = 1, \ldots, m^*$ need to be generated only once from $k = M-1$ to $k = 1$ before we generate the diffusion bridge samples. The function $f_k(\cdot)$ for all $k = 1, \ldots, M-2$, are estimated in this process; thus the extra computational burden for calculating $f_k(\cdot)$ is limited. In addition, this stage serves as a general guidance for resampling. A highly 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; thus the required number of backward pilots m^* need not be large.

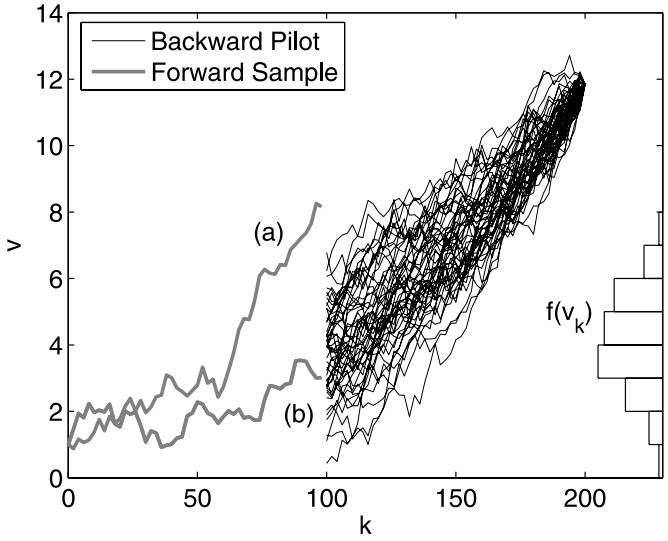


Figure 2. Using the backward pilots to obtain the resampling priority score.

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

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

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

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

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

$$\begin{aligned}
\mu_k^{(j)} &= \big(\boldsymbol{\Sigma}_k^{*(j)}\big)^{-1}\big(u_{k+1}^{(j)} - \mathbf{b}\big(u_k^{*(j)}, \boldsymbol{\theta}\big)\delta \\
&\quad + \mathbf{H}\big(u_k^{*(j)}; \boldsymbol{\theta}\big)u_k^{*(j)}\delta\big), \\
\boldsymbol{\Sigma}_k^{(j)} &= \big(\boldsymbol{\Sigma}_k^{*(j)}\big)^{-1}\mathbf{A}\big(u_k^{*(j)}; \boldsymbol{\theta}\big)\mathbf{A}^T\big(u_k^{*(j)}; \boldsymbol{\theta}\big)\big(\boldsymbol{\Sigma}_k^{*(j)}\big)^{-T}\delta, \\
\boldsymbol{\Sigma}_k^{*(j)} &= \mathbf{I}_d + \mathbf{H}\big(u_k^{*(j)}; \boldsymbol{\theta}\big)\delta.
\end{aligned}$$

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

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

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

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

including the transition probability (as a function of v_k)

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

5. In step 2(c) of the algorithm, although the function $\widehat{f}_k(\cdot)$ estimated using the 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 the 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, and w_t is Brownian motion. This diffusion process actually exhibits certain jump phenomena. 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 toward $\theta + \pi + 2k\pi$. Figure 3 shows a realization of the process with parameter $\theta = \pi$.

3.1.1 Estimation of the Transition Density With Parameter $\theta = \pi$. We first consider the generation of the diffusion bridges with two fixed endpoints at V_0 and V_{30}, using $M = 400$ intermediate time points. To simplify notation, we use SMC-0 to denote Durham and Gallant (2002)'s sampler (9) without resampling, and 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.

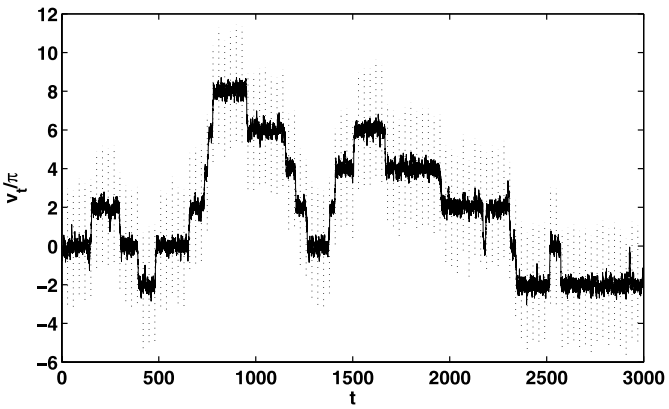


Figure 3. A realization of the process with sine drift coefficients following equation (14), with $\theta = \pi$. The vertical lines present observations with time interval $\Delta = 30$.

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 from these 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 samples 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 $\mathbf{v}^{(j)}$, $j = 1, \ldots, m$, are generated independently, we have

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

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

Thus 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 endpoints V_0 and V_{30}, we repeat the estimation 100 times independently and calculate

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

as the performance measurement, where $\widehat{P}^{*(i)}(V_{30} \mid V_0, \theta = \pi)$ is the ith 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 using Beskos et al. (2006)'s exact sampling method with $m = 10{,}000{,}000$ Monte Carlo samples. With roughly equal computational time, we used $m = 3500$ samples for SMC-0 and $m = 1000$ samples for SMC-1. Note that the sampling distributions of SMC-0 do not depend on the parameter, and that the samples can be linearly transformed to meet different fixed endpoints, although the weight calculation depends on θ and the endpoints. Thus the computational time for SMC-0 involves only the time to evaluate $\prod_{k=1}^{M} P^*(v_k^{(j)} \mid v_{k-1}^{(j)}; \theta = \pi)$.

Table 1 reports the ratio of $\mathrm{RMSE}(V_0, V_{30})$ of SMC-1 to $\mathrm{RMSE}(V_0, V_{30})$ of SMC-0. It can be seen that for most of the

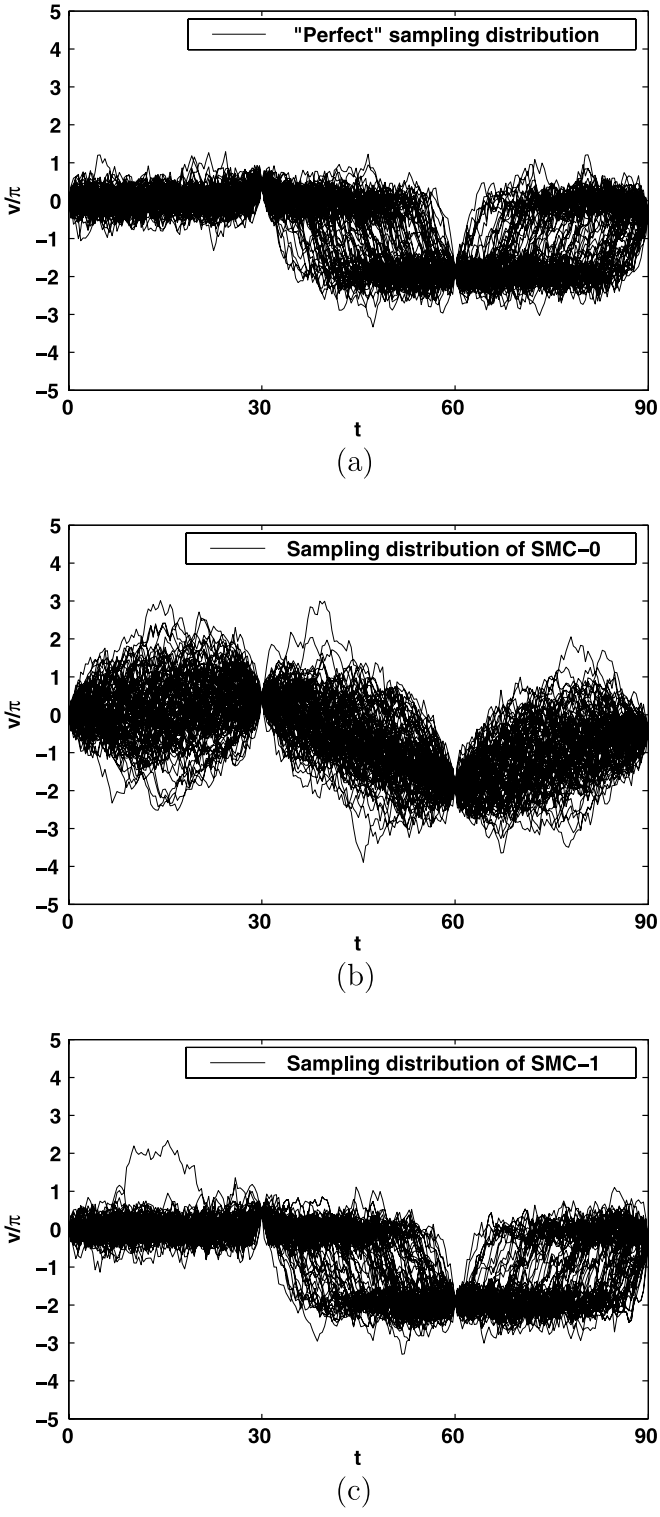


Figure 4. Illustration of bridge samples of the sine drift process generated using different sampling methods with $M = 400$ intermediate points between two consecutive observations. The parameter is $\theta = \pi$, and the observations are $(V_0, V_{30}, V_{60}, V_{90}) = (0, 1.49, -5.91, -1.17)$. (a): 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.

Table 1. The ratio of RMSE(V_0, V_{30}) for SMC-1 to RMSE(V_0, V_{30}) for SMC-0 for estimating $\log P(V_{30} \mid V_0; \theta = \pi)$ for the diffusion process with sine drift

	V_0										
$\frac{\text{RMSE}_{\text{SMC-1}}(V_0,V_{30})}{\text{RMSE}_{\text{SMC-0}}(V_0,V_{30})}$	-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

NOTE: The sample sizes ($m = 3500$ for SMC-0 and $m = 1000$ for SMC-1) are controlled, so the two methods have similar computational times.

(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 occur only 0.1% of the time. Tables 2 and 3 report the RMSE(V_0, V_{30}) of SMC-0 and SMC-1, respectively. They show that the RMSE(V_0, V_{30}) of SMC-0 decreases slowly as $|V_{30} - V_0|$ increases, and that the RMSE(V_0, V_{30}) of SMC-1 maintains a relatively small value when V_0 and V_{30} are close, but increases rapidly with increasing distance between V_0 and V_{30} when $|V_{30} - V_0| > 4\pi$. It seems that when there are several unidirectional jumps between the two endpoints in a short time interval, the most likely paths are those close to the straight line between the two endpoints. In this case, the sampling distribution used by Durham and Gallant (2002) (SMC-0) guides the sampling more forcefully to reach the far-away target endpoint. In contrast, the resampling approach (SMC-1) might require larger Monte Carlo sample sizes of the backward pilots to obtain "good" resampling priority scores, because the backward pilots are generated without considering the location of v_0. To illustrate the variability of the estimators, Figure 5 shows a boxplot of the estimation er-

Table 2. 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 $m = 3500$

	V_0										
$\text{RMSE}_{\text{SMC-0}}(V_0, V_{30})$	-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 3. 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 $m = 1000$

	V_0										
$\text{RMSE}_{\text{SMC-1}}(V_0, V_{30})$	-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

rors of the log transition density for a selected set of (V_0, V_{30}) combinations. It can be seen that the error distribution is relatively symmetric and well behaved; thus the RMSE accurately reflects the performance of estimation methods. The log transition probability estimated by SMC-1 with $m = 1000$ samples is plotted in Figure 6.

Figure 7 shows the sampled diffusion bridges for selected pairs of (V_0, V_{30}) using the SMC-1 method. It is seen that the process is stable around $0, 2\pi, 4\pi, \ldots$ and transient around $\pi, 3\pi$, etc., and exhibits jump behavior after entering the transition zone, spending a very short time inside the transit zone. Thus the process tends to remain in the stable zone if not required to move to another zone [panel (a)]. If the endpoint 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 [(b), (d), and (f)]. Panel (b) also shows that the process is also able to jump 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 but different zones, then the jump will occur within the time period and the timing will be almost uniform, except at the beginning and the end [(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

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

In what follows, 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. Thus the observations are observed at $t = 0, 30, 60, \ldots, 3000$. The paths are simulated using Euler 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.

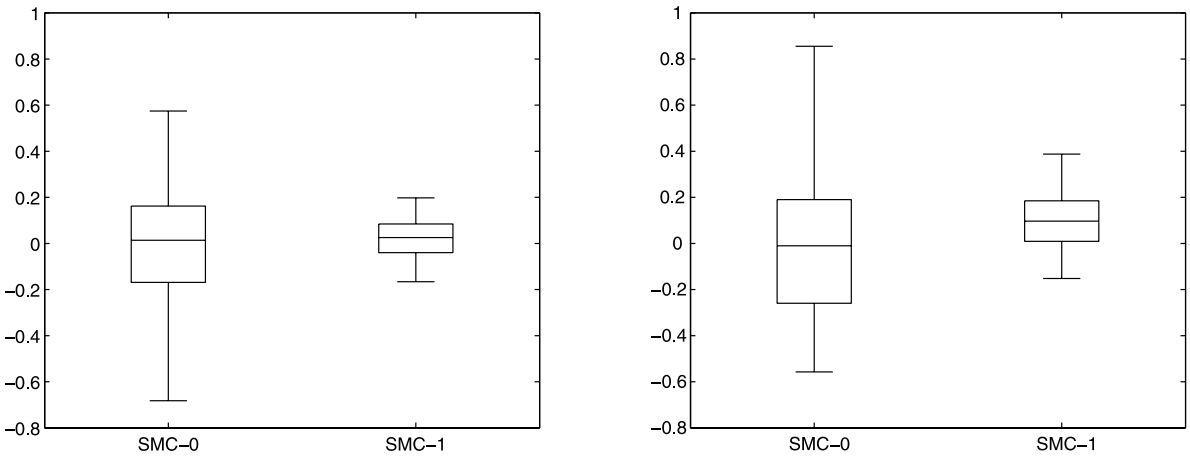
We use the following measure of efficiency:

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

Again, we obtain the "true" value of $L_i(\theta)$ via the exact sampling method with a Monte Carlo sample size of $m = 10{,}000{,}000$.

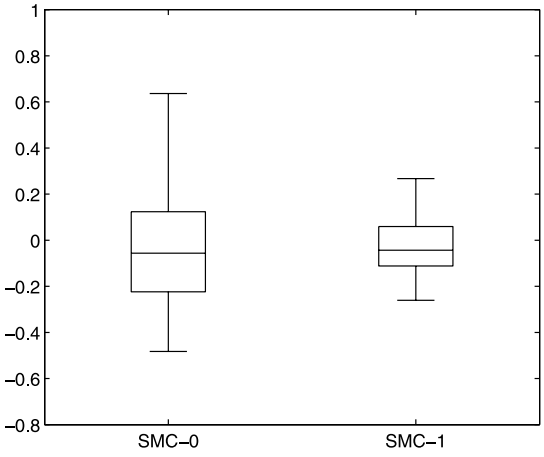
We report the average RMSE(θ) of the 100 simulated paths for different methods in Table 4. For each method, its Monte Carlo sample size m is chosen so that the methods take approximately the same CPU time. From the table we can see that SMC-1 performs the best for estimating $L(\theta)$ of all θ (the 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 8 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

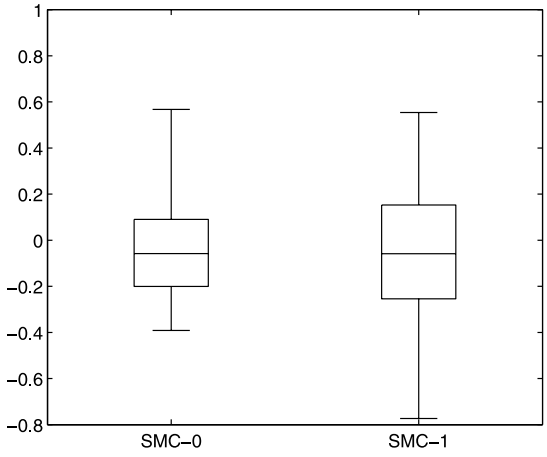


(a) $V_0 = 0, \quad V_{30} = 0$

(b) $V_0 = 0, \quad V_{30} = \pi$



(c) $V_0 = 0, \quad V_{30} = 2\pi$



(d) $V_0 = 0, \quad V_{30} = 4\pi$

Figure 5. Boxplots of estimation errors in 100 independent estimations of $\log P(V_{30} \mid V_0, \theta = \pi)$ for the sine-drift example. The performance of SMC-0 and SMC-1 is compared. V_{30} is 0 in (a), π in (b), 2π in (c), and 4π in (d).

smooth likelihood function. Further research on this direction is needed.

3.2 Example 2

We next consider the jump diffusion process (Merton 1976; Cox, Ross, and Rubinstein 1979; Aït-Sahalia 2004)

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

which is often used for modeling stock prices. We set $\alpha = 0.08$, $\sigma = 0.2$, $\lambda = 5$, and z_t as a Poisson process, with λ the mean number of arrivals per unit time. In addition, when an Poisson event occurs, z_t produces a jump of size y that 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 9.

For a small time interval δ, let

$$\overline{v}_{t+\delta} = \begin{cases} \overline{v}_t + \left(\alpha - \lambda\kappa - \dfrac{\sigma^2}{2}\right)\delta + w_{t+\delta} - w_t + z_{t+\delta} - z_t \\ \quad \text{if} \le 1 \text{ jumps occur in } [t, t+\delta) \\ \overline{v}_t + \left(\alpha - \lambda\kappa - \dfrac{\sigma^2}{2}\right)\delta + w_{t+\delta} - w_t \\ \quad \text{if} \ge 2 \text{ jumps occur in } [t, t+\delta). \end{cases}$$

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

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

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. Thus for this process,

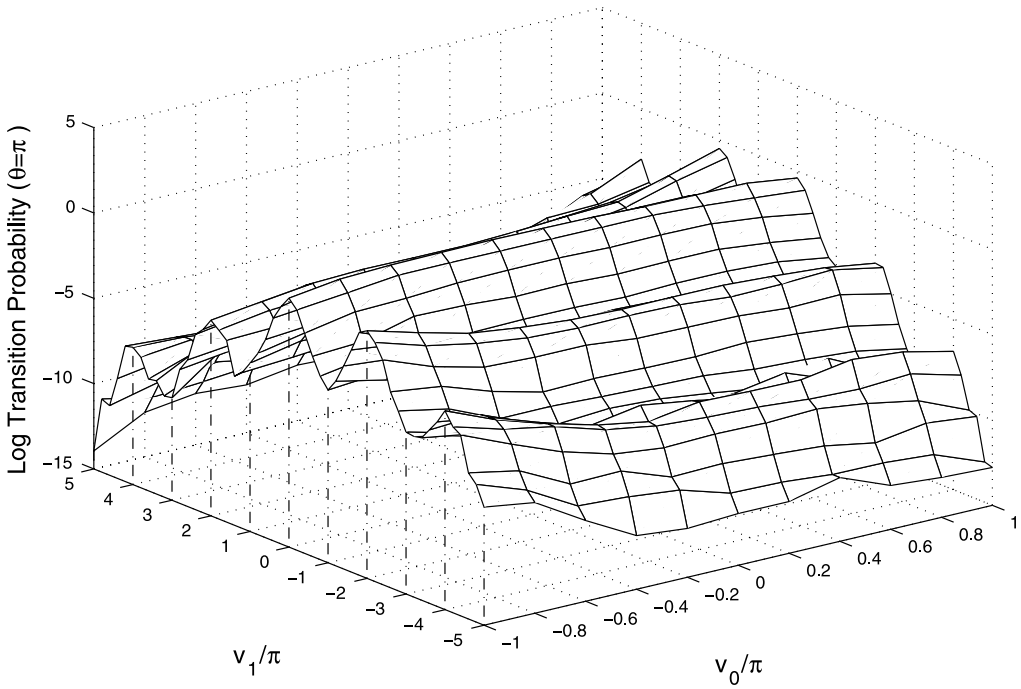


Figure 6. The log transition probability of the sine drift process estimated using SMC-1. The sample size is $m = 1000$.

we can approximate $P(v_k \mid v_{k-1})$ by

$$P^*(v_k \mid v_{k-1}) \sim \begin{cases} \mathcal{N}\left(v_{k-1} + \left(\alpha - \lambda\kappa - \dfrac{\sigma^2}{2}\right)\delta, \sigma^2\delta + \sigma_y^2\right) \\ \quad \text{with probability } 1 - \lambda\delta \\ \mathcal{N}\left(v_{k-1} + \left(\alpha - \lambda\kappa - \dfrac{\sigma^2}{2}\right)\delta, \sigma^2\delta\right) \\ \quad \text{with probability } \lambda\delta. \end{cases} \tag{16}$$

3.2.1 Transition Density Estimation. In this example, the transition probability density $P(V_\Delta \mid V_0)$ depends only on the time interval Δ and the difference, $V_\Delta - V_0$, between the two endpoints. We fixed $V_0 = 0$ and considered the effect of different durations of time period $\Delta = i/36$, $i = 1, \ldots, 9$ and different ending points V_Δ. To accommodate the different time periods, 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

$$r_k(v_k \mid v_{k-1}) = P^*(v_k \mid v_{k-1}), \tag{17}$$

and Durham and Gallant (2002)'s sampler as

$$r_k(v_k \mid v_{k-1}) \sim \begin{cases} \mathcal{N}\left(v_{k-1} + \dfrac{v_M - v_{k-1}}{M-k+1}, \dfrac{M-k}{M-k+1}\sigma^2\delta + \sigma_y^2\right) \\ \quad \text{with probability } 1 - \lambda\delta \\ \mathcal{N}\left(v_{k-1} + \dfrac{v_M - v_{k-1}}{M-k+1}, \dfrac{M-k}{M-k+1}\sigma^2\delta\right) \\ \quad \text{with probability } \lambda\delta. \end{cases} \tag{18}$$

It can be verified that both the *extended* Pedersen (1995) sampler and the *extended* Durham and Gallant (2002) sampler satisfy the consistent condition $\mathrm{Var}_Q(w) < \infty$. For this example, our numerical experiment shows that sampler (17) outperforms sampler (18) with the same computational time. Thus we use (17) as a 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 $[0.04l, 0.04(l+1))$, $l = 0, \pm1, \pm2, \ldots$.

The sampling distribution for generating the backward pilots is

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

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

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

For $\Delta = 1/12$, we plot 100 bridge samples generated by the "perfect" sampling distribution, SMC-0, and SMC-1 in Figure 10 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 with that under SMC-0. Thus SMC-1 is more efficient.

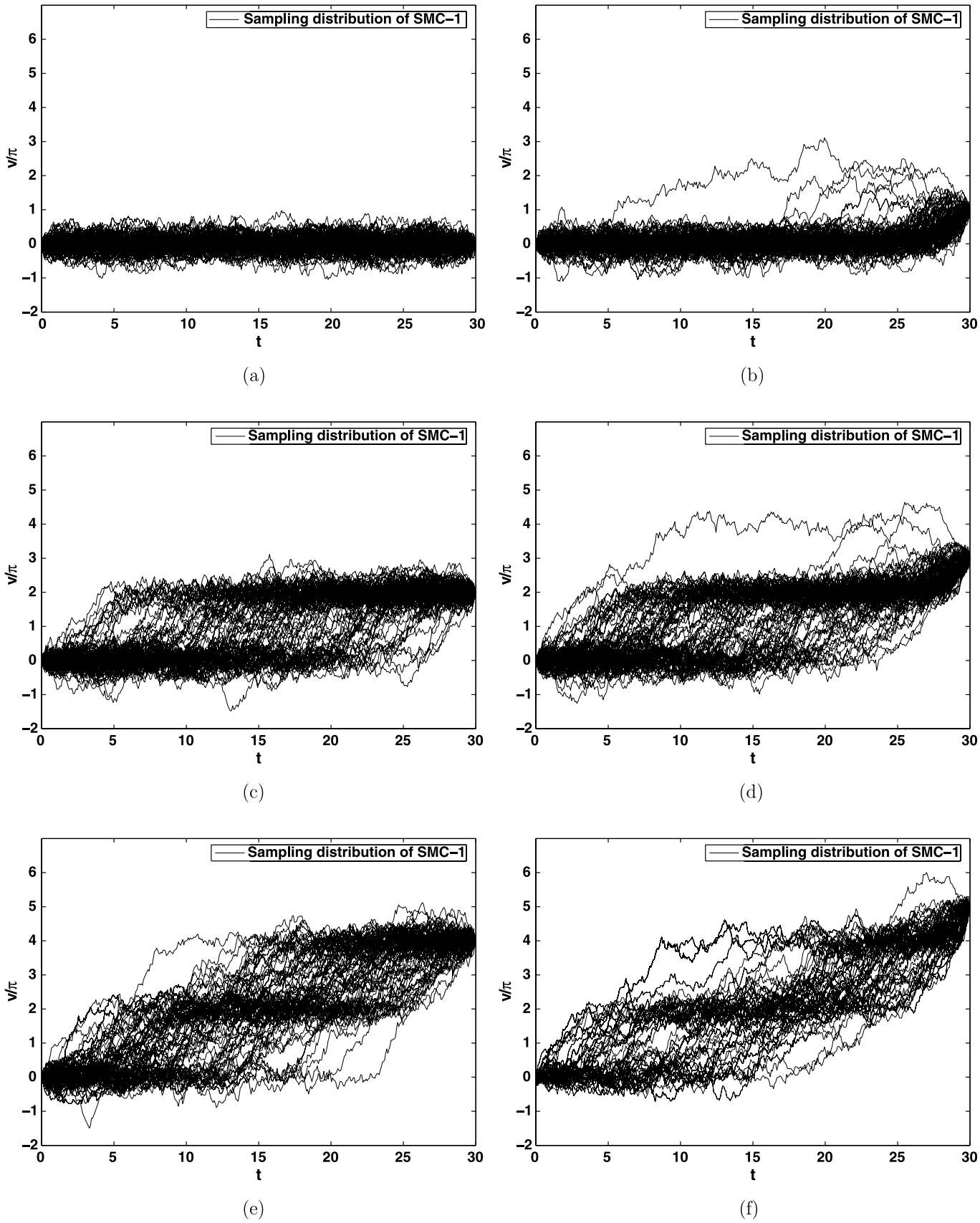


Figure 7. Illustration of bridge samples of the sine drift process generated by SMC-1 for $\theta = \pi$, $V_0 = 0$ and different values of V_{30}. The resampling step is performed every 20 steps when generating the bridge samples. $m^* = 300$ backward pilots are generated to estimate the resampling priority scores.

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 $\theta = \pi$

RMSE(θ)	Exact sampling	SMC-0	SMC-1
m	80,000	3500	1000
$\theta = 0.0\pi$	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

NOTE: Row 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.

For transition density estimation, we obtained

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

over 100 independent trials, for various Δ's and different V_Δ's. 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 $m = 5000$ and $m = 2000$ Monte Carlo samples,

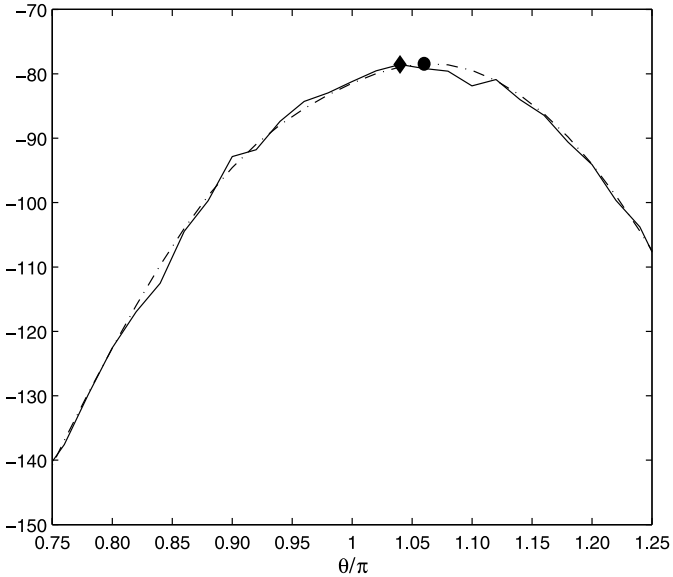


Figure 8. Estimated log-likelihood function using SMC-1 (solid line) and the "true" log-likelihood function (dashed line), with a grid points of every 0.2π, based on a simulated sample path observed at $t = 0, 30, 60, \ldots, 3000$ (101 observations). The corresponding MLEs are shown as the diamond ($\hat{\theta} = 1.04$) for SMC-1 and the circle ($\hat{\theta} = 1.06$) under the "true" log-likelihood function.

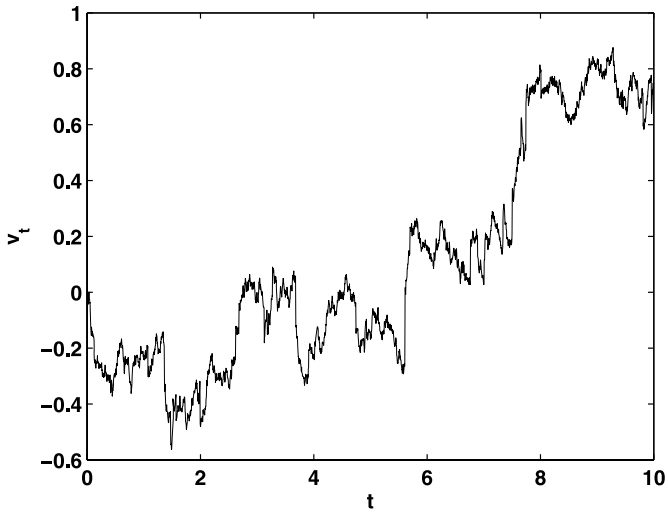


Figure 9. A realization of the jump diffusion process following eq. (15) with $\alpha = 0.08$, $\sigma = 0.2$, $\lambda = 5$, $\mu_y = 0$, and $\sigma_y = 0.1$.

respectively. We plot RMSE(Δ, V_Δ) for $\Delta = 1/36$, $\Delta = 3/36$, and $\Delta = 9/36$ in Figure 11(b), (d), and (f), along with the true value of the transition density $P(V_\Delta \mid V_0 = 0)$ in (a), (c), and (e). Figure 12 also shows the boxplot of estimation errors of the transition density for a selected sets of $V_{1/4}$ values, in 100 repeated experiments. Again the error distribution is relatively well behaved.

Table 5 reports the overall performance measure

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

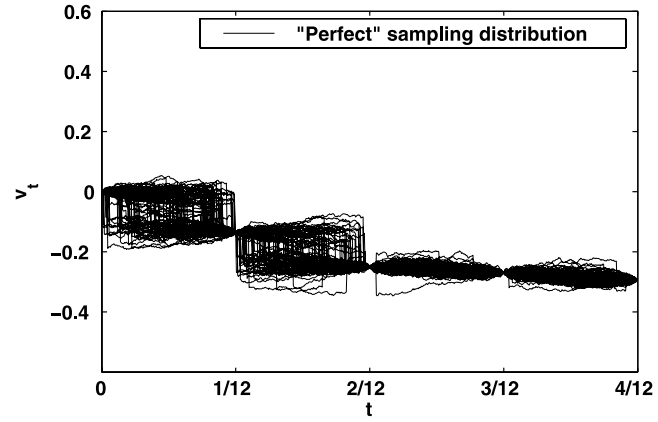
while controlling the CPU time to be roughly the same for different methods. Clearly, SMC-1 is more efficient.

3.2.2 Estimation of the Realized Volatility. Another interesting use of diffusion bridge samples is in estimating the realized volatility (Hull and White 1987; Barndorff-Nielsen and Shephard 2002; Zhang, Mykland, and Aït-Sahalia 2005) conditional on the two endpoints. 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

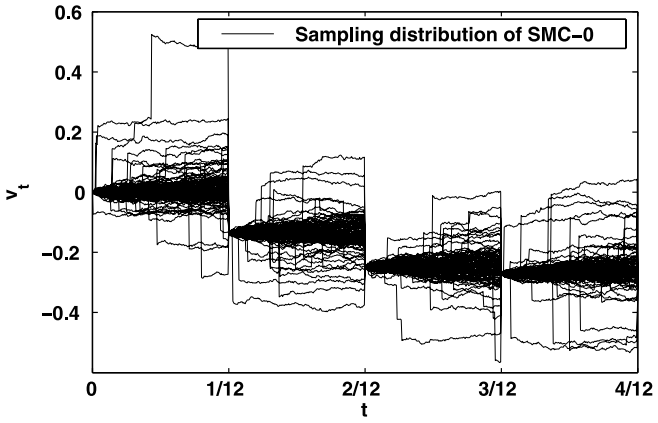
$$\widehat{I}(\Delta, V_\Delta) = \frac{\sum_{j=1}^{m} w^{(j)} \sum_{s=1}^{M} (v_s^{(j)} - v_{s-1}^{(j)})^2}{\sum_{j=1}^{m} w^{(j)}}. \tag{19}$$

In fact, there is a better approach to resampling the priority score for this specific inference problem. The "optimal" resampling priority score developed by (11) was designed only to generate the "best" bridge samples and to minimize the variance of weight, $\mathrm{Var}_Q\, w$.

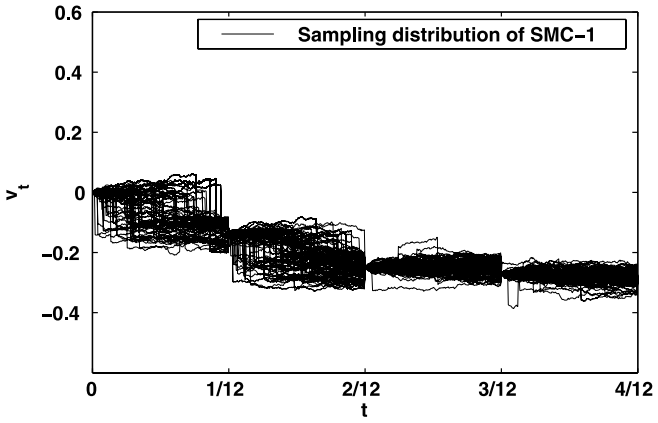
At step k ($k \le M - 2$), when constructing the priority score $\beta_k^{(j)}$ for estimating $I(\Delta, V_\Delta)$, we consider minimizing the variance of $w^{(j)} (\sum_{s=1}^{M} (v_s^{(j)} - v_{s-1}^{(j)})^2)^{1/2}$ instead of minimizing the variance of $w^{(j)} \sum_{s=1}^{M} (v_s^{(j)} - v_{s-1}^{(j)})^2$, which is the term $\mathrm{Var}_Q(wh)$ [or, equivelently, $\mathrm{Var}_Q(\overline{w}h)$] in the MSE approximation (7). Again, given the forward sampling distributions $r_s(v_s \mid v_{s-1}, v_M; \theta)$, $s = k+1, \ldots, M-1$, without considering



(a)



(b)



(c)

Figure 10. 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.

the effect of future resampling after step k, the problem is equivalent to finding β that minimizes

$$
\begin{aligned}
&E\left[w^2 \sum_{s=1}^{M} (v_s - v_{s-1})^2 \,\middle|\, V_0, V_\Delta \right] \\
&\quad = E\left[\left(\frac{\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)}{\beta_k Q_k(\mathbf{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \right)^2 \right. \\
&\qquad \left. \times \sum_{s=1}^{M} (v_s - v_{s-1})^2 \,\middle|\, V_0, V_\Delta \right] \\
&\quad = \int \frac{[\prod_{s=1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\beta_k Q_k(\mathbf{v}_k) \prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \\
&\qquad \times \sum_{s=1}^{M} (v_s - v_{s-1})^2 \, dv_1 \cdots dv_{M-1} \\
&\quad = \int \frac{[\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)]^2}{\beta_k Q_k(\mathbf{v}_k)} \sum_{s=1}^{k} (v_s - v_{s-1})^2 \\
&\qquad \times \int \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1} \, dv_1 \cdots dv_k \\
&\qquad + \int \frac{[\prod_{s=1}^{k} P^*(v_s \mid v_{s-1}; \theta)]^2}{\beta_k Q_k(\mathbf{v}_k)} \int \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \\
&\qquad \times \sum_{s=k+1}^{M} (v_s - v_{s-1})^2 \, dv_{k+1} \cdots dv_{M-1} \, dv_1 \cdots dv_k.
\end{aligned}
$$

Thus $\mathrm{Var}[w\sqrt{\sum_{s=1}^{M}(v_s - v_{s-1})^2}]$ is minimized when

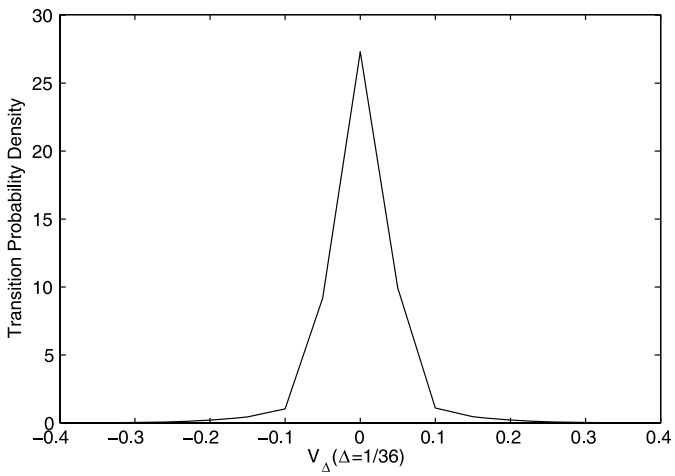
$$
\begin{aligned}
\beta_k = w_k \Bigg[& \sum_{s=1}^{k} (v_s - v_{s-1})^2 \\
& \times \int \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1} \\
& + \int \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1}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

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

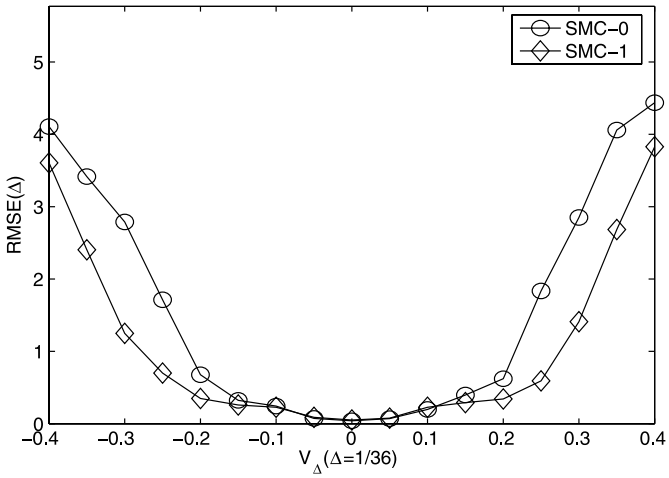
$$
\widehat{f}_k^{(1)}(v_k; \theta) \simeq \int \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \, dv_{k+1} \cdots dv_{M-1}
$$

and

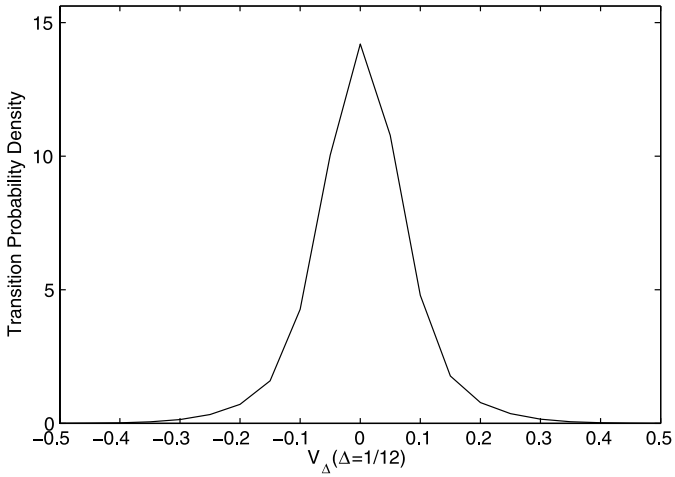
$$
\begin{aligned}
\widehat{f}_k^{(2)}(v_k; \theta) \simeq \int & \frac{[\prod_{s=k+1}^{M} P^*(v_s \mid v_{s-1}; \theta)]^2}{\prod_{s=k+1}^{M-1} r_s(v_s \mid v_{s-1}, v_M; \theta)} \\
& \times \sum_{s=k+1}^{M} (v_s - v_{s-1}) \, dv_{k+1} \cdots dv_{M-1}.
\end{aligned}
$$



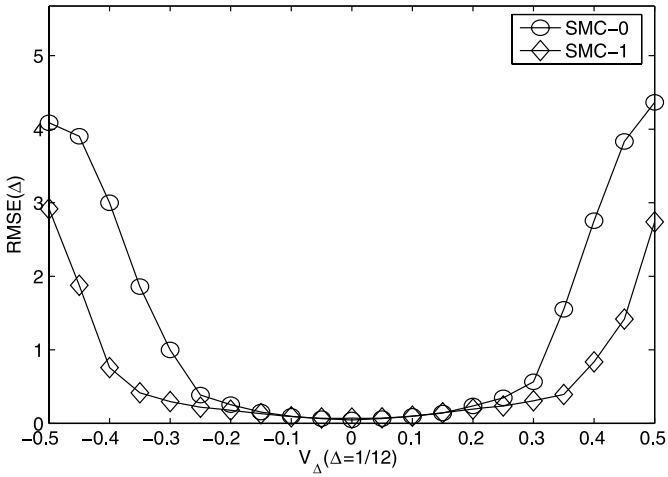
(a)



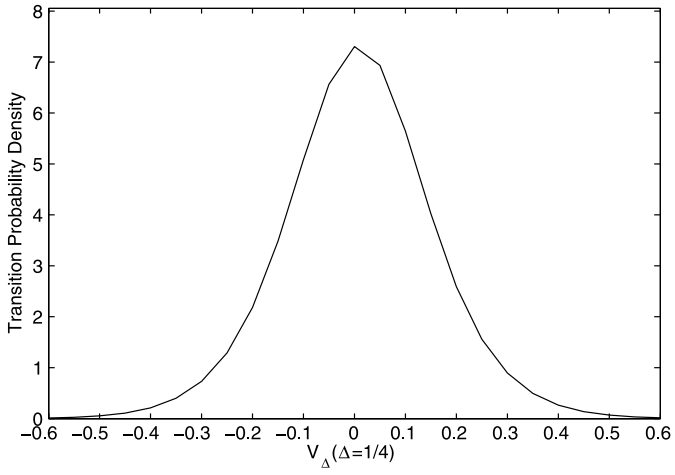
(b)



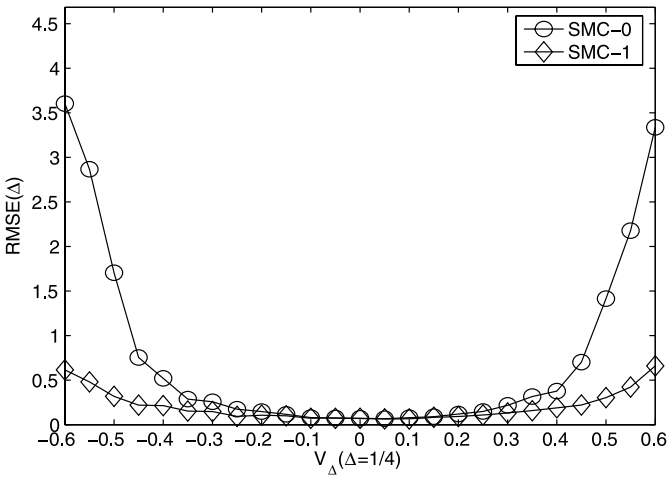
(c)



(d)

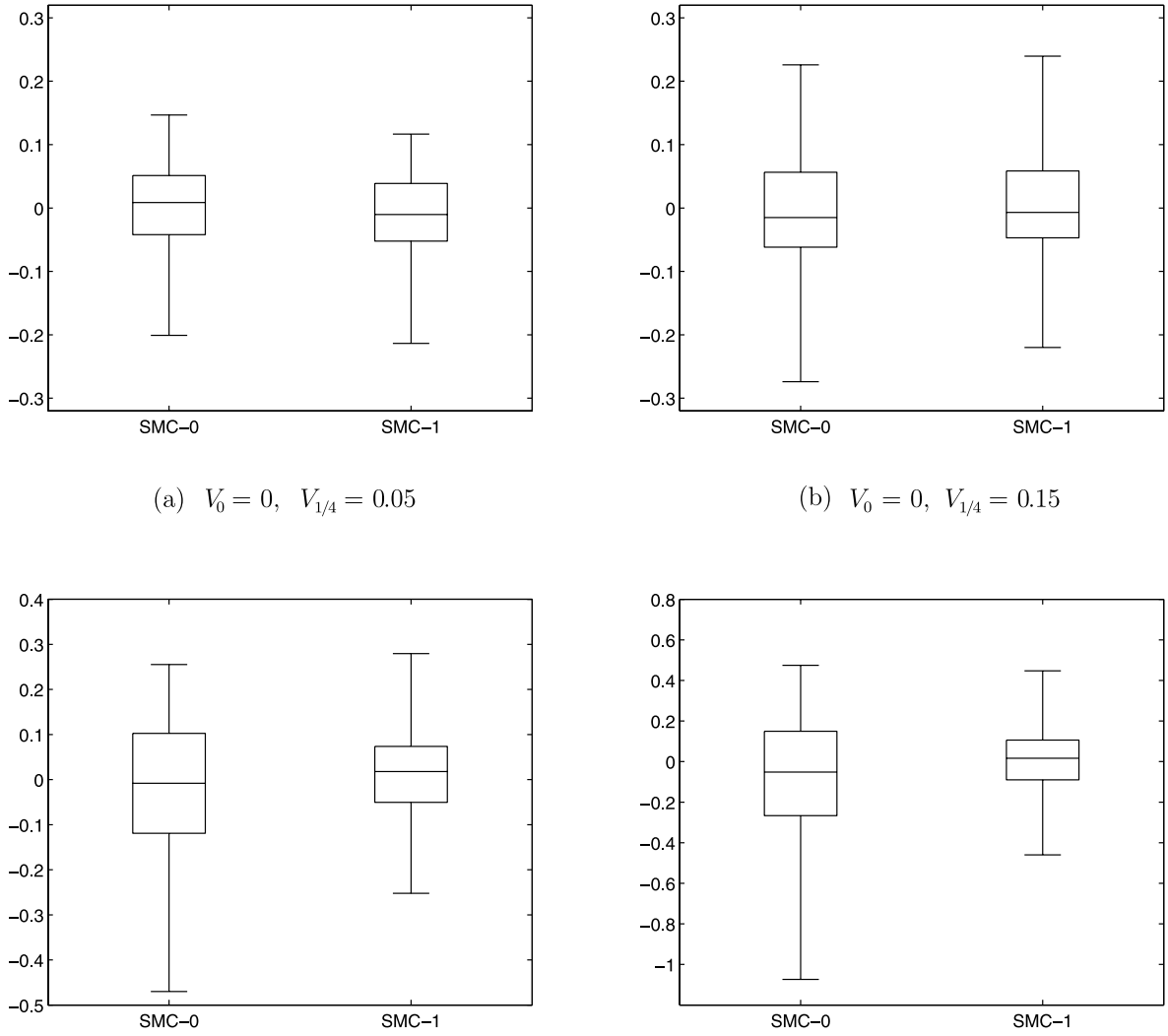


(e)



(f)

Figure 11. (a), (c), and (e) True transition density $P(V_\Delta \mid V_0 = 0)$ for the jump diffusion process with observation time intervals $\Delta = 1/36$, $\Delta = 1/12$, and $\Delta = 1/4$, respectively. (b), (d), and (f) The corresponding RMSE(Δ, V_Δ) of estimating the transition densities using SMC-0 with $m = 5000$ samples and SMC-1 with $m = 2000$ samples.



(a) $V_0 = 0, \quad V_{1/4} = 0.05$ (b) $V_0 = 0, \quad V_{1/4} = 0.15$

(c) $V_0 = 0, \quad V_{1/4} = 0.25$ (d) $V_0 = 0, \quad V_{1/4} = 0.35$

Figure 12. Boxplots of estimation errors of $\log P(V_{\Delta=1/4} \mid V_0 = 0)$ for the jump process, in 100 independent repeated experiments. We compared the performance of SMC-0 and SMC-1. $V_{1/4}$ is 0, 0.1, 0.2, and 0.35 in (a), (b), (c), and (d), respectively.

Table 5. RMSE(Δ) of using SMC-0 and SMC-1 to estimate the log-transition probability of the jump diffusion process

RMSE(Δ) of log-likelihood	SMC-0	SMC-1
m	5000	2000
$\Delta = 1/36$	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

NOTE: Row 2 is the Monte Carlo sample size (m), and the last row reports the average CPU time of each evaluation of the log transition density.

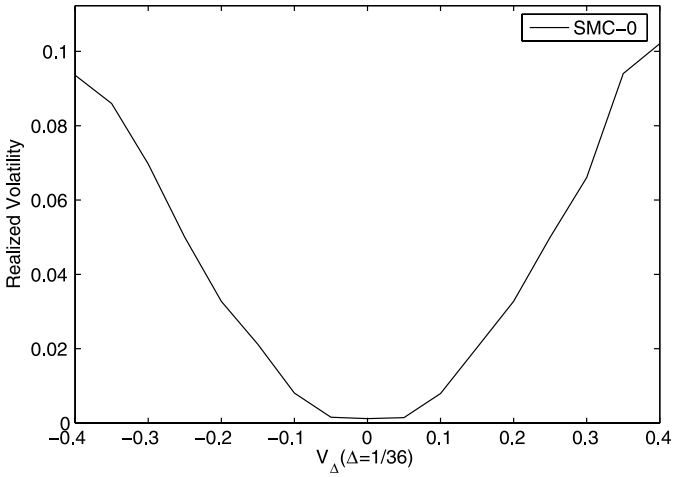
Then the priority score is set as

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

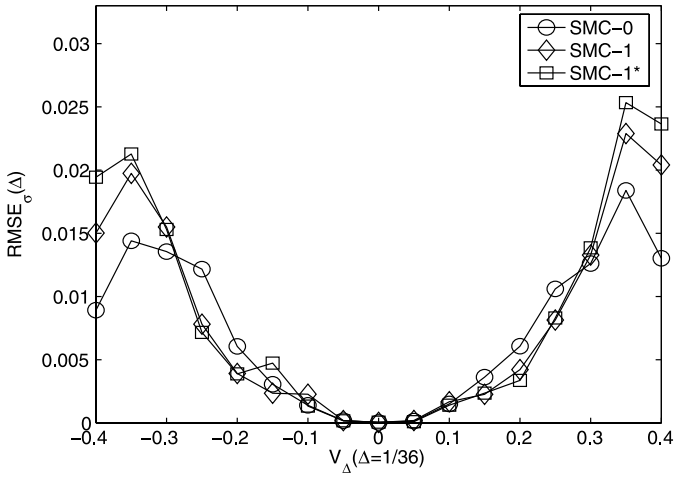
The SMC that uses the resampling priority score (20) is designated SMC-1*, whereas SMC-1 uses (11) as the resampling priority score. The other settings in SMC-1* are the same as in SMC-1. For a performance comparison, define

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

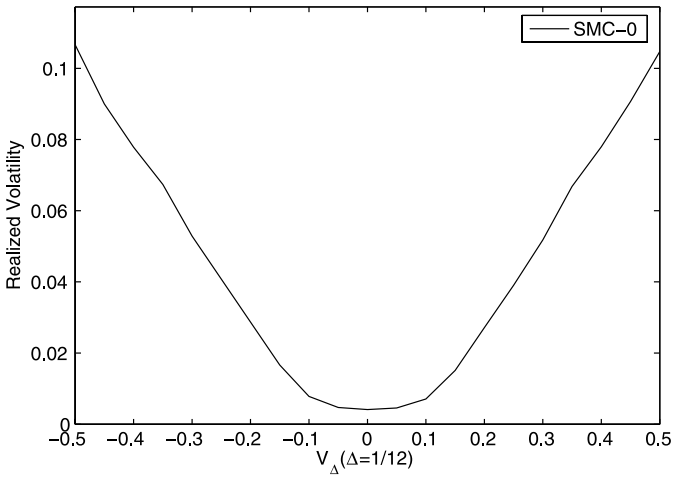
over 100 independent estimates as the measurement of estimate accuracy, where the "true" value of $I(\Delta, V_\Delta) = E(\sum_{s=1}^{M} (v_s - v_{s-1})^2 \mid V_0, V_\Delta)$ is obtained by using SMC-0 with $m = 500{,}000$ Monte Carlo samples. Values of $\mathrm{RMSE}_\sigma(\Delta, V_\Delta)$ for $\Delta = 1/36$, $\Delta = 3/36$, and $\Delta = 9/36$ are plotted in Figure 13. Figure 13 also shows the "true" value of $E(\sum_{s=1}^{M} (v_s - v_{s-1})^2 \mid V_0, V_\Delta)$. It



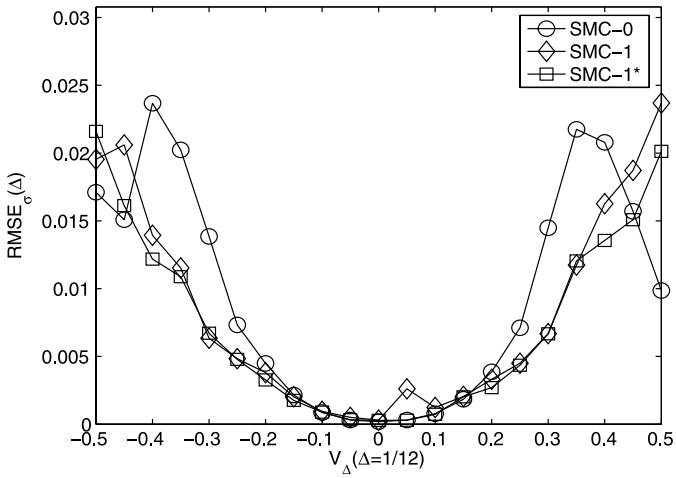
(a)



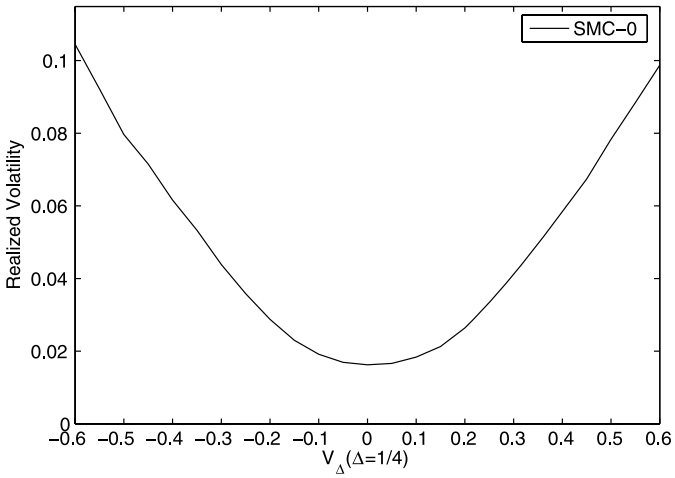
(b)



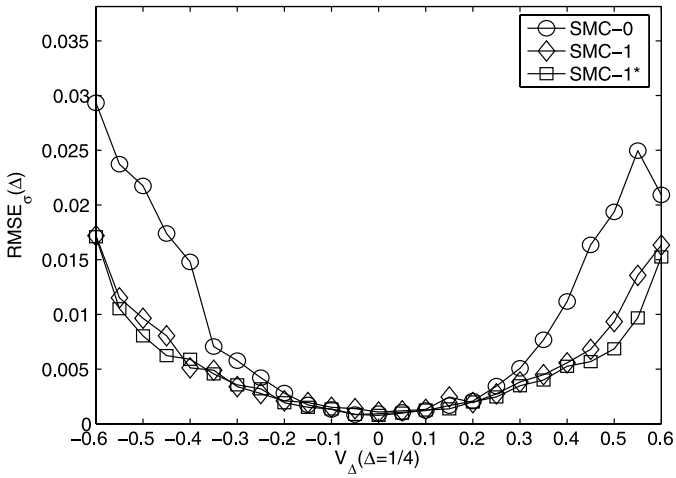
(c)



(d)



(e)



(f)

Figure 13. (a), (c), and (e) Plots of the "true" values of $E(\sum_{s=1}^{M}(v_s - v_{s-1})^2 \mid V_0 = 0, V_\Delta)$ of the jump diffusion process with observation time intervals $\Delta = 1/36$, $\Delta = 1/12$, and $\Delta = 1/4$, respectively. (b), (d), and (f) show the corresponding $\mathrm{RMSE}_\sigma(\Delta, V_\Delta)$ using SMC-0 with $m = 5000$ samples, SMC-1 with $m = 2000$ samples, and SMC-1* with $m = 2000$ 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.

Table 6. $\text{RMSE}_\sigma(\Delta)$'s 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

$\text{RMSE}_\sigma(\Delta)$ ($\times 10^2$)	SMC-0	SMC-1	SMC-1*
m	5000	2000	2000
$\Delta = 1/36$	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

NOTE: Row 2 is the Monte Carlo sample sizes (m), and the last row reports the average CPU time of each evaluation.

can be seen that the SMC methods with backward pilot-guided resampling perform well, especially for longer time periods Δ and when the difference between the two endpoints $|V_\Delta - V_0|$ is large.

Table 6 reports the overall performance measurement, defined as

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

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 article we have proposed a resampling scheme under the SMC framework for the purpose of generating samples of diffusion bridges that connect the two observed ends of the underlying diffusion process. This resampling scheme can be easily combined with various sampling methods, including the sampling methods of Pedersen (1995) and Durham and Gallant (2002). The key idea is to use generated backward pilots to estimate the optimal resampling priority scores for adjustment of the sampling distribution in the intermediate steps. The additional computational complexity for computing the "optimal" priority scores is limited.

Although the two illustrating examples are on one-dimensional diffusion processes, the algorithm is designed for multidimensional processes. In such cases, the use of backward pilot samples is still effective for exploring the intermediate space through which the bridge paths are likely to pass. Certainly, a larger number of such pilots is needed. However, the extra computational burden remain limited, because the backward pilots needs only be generated once and off-line, and a very accurate estimation of the optimal priority score is not needed to achieve effective resampling. Thus the requirement for the number of backward pilots needed is not extremely demanding. Nevertheless, dealing with the multidimensional processes is much more difficult. If the process can be decomposed or transformed so that there is a sequential structure between the dimensions, then samples of the (lower-dimensional) marginal and conditional process can be generated sequentially. For example, let $V_t = (V_{t1}, \ldots, V_{td})$. If for each k, the kth draft coefficient $b_k(V_t; \boldsymbol{\theta})$ in (1) depends only on $(V_{t1}, \ldots, V_{tk})$, the first k component of V_t, and the diffusion matrix $\mathbf{A}(V_t, \boldsymbol{\theta})$ is of lower triangle, with each nonzero element $a_{i,j}(V_t; \boldsymbol{\theta})$ depending only on $(V_{t1}, \ldots, V_{tj})$, then the sample paths can be generated for V_{t1} following

$$dV_{t1} = b_1(V_{t1}; \boldsymbol{\theta})\, dt + a_{1,1}(V_{t1}; \boldsymbol{\theta})\, dW_{t1}.$$

Then, conditional on each sample path $V_{t1}^{(j)}$, the second component, V_{t2}, can be generated following

$$dV_{t2} = b_2\big(V_{t1}^{(j)}, V_{t2}; \boldsymbol{\theta}\big)\, dt + a_{1,2}\big(V_{t1}^{(j)}, V_{t2}; \boldsymbol{\theta}\big)\, dW_{t1}^{(j)} + a_{2,2}\big(V_{t1}^{(j)}, V_{t2}; \boldsymbol{\theta}\big)\, dW_{t2}.$$

The rest of the components can be generated similarly. Further research on this approach is needed.

Also note that the "optimal" priority scores might be different for different objectives of generating diffusion bridge samples. In cases of parameter estimations, likelihood function estimates often are expected to be smooth and continuous (Pitt 2002). More research is needed to properly adjust our algorithm to meet such a requirement.

[Received January 2009. Revised February 2010.]

REFERENCES

Aït-Sahalia, Y. (2004), "Disentangling Diffusion From Jumps," *Journal of Financial Economics*, 74, 487–528. [829]

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, Ser. B*, 64, 253–280. [832]

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, Ser. B*, 68, 333–382. [820,825,826,828]

Beskos, A., Roberts, G., Stuart, A., and Voss, J. (2008), "MCMC Methods for Diffusion Bridges," technical report, University of Warwick. [821]

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. [820,821]

Cox, J., Ross, S., and Rubinstein, M. (1979), "Option Pricing: A Simplified Approach," *Journal of Financial Economics*, 7, 229–263. [829]

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. [820-823, 825-827,830,837]

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959–993. [820-822]

Eraker, B. (2001), "MCMC Analysis of Diffusion Models With Application to Finance," *Journal of Business & Economic Statistics*, 19, 177–191. [820]

Gilks, W., Richardson, S., and Spiegelhalter, D. (1995), *Markov Chain Monte Carlo in Practice*, New York: Chapman & Hall. [821]

Hull, J., and White, A. (1987), "The Pricing of Options on Assets With Stochastic Volatilities," *Journal of Finance*, 42, 281–300. [832]

Kloeden, P., and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Berlin: Springer-Verlag. [820]

Kong, A., Liu, J., and Wong, W. (1994), "Sequential Imputations and Bayesian Missing Data Problems," *Journal of the American Statistical Association*, 89, 278–288. [822]

Liu, J., and Chen, R. (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93, 1032–1044. [822-824]

Liu, J. S. (1996), "Metropolized Independent Sampling With Comparisons to Rejection Sampling and Importance Sampling," *Statistics and Compustion*, 6, 113–119. [822]

——— (2001), *Monte Carlo Strategies in Scientific Computing*, New York: Springer. [821,822]

Marshall, A. (1956), "The Use of Multi-Stage Sampling Schemes in Monte Carlo Computations," in *Symposium on Monte Carlo Methods*, ed. M. Meyer, New York: Wiley, pp. 123–140. [821]

Merton, R. C. (1976), "Option Pricing When Underlyinh Stock Returns Are Discontinous," *Journal of Financial Economics*, 3, 125–144. [829]

Pedersen, A. R. (1995), "Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 257–279. [820-822,830,833,837]

Pitt, M. K. (2002), "Smooth Particle Filters for Likelihood and Maximisation," technical report, University of Warwick. [828,837]

Robert, R., and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer. [821]

Roberts, G. O., and Stramer, O. (2001), "On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis–Hastings Algorithm," *Biometrika*, 88, 603–621. [820,821]

Shoji, I., and Ozaki, T. (1998), "Estimation for Nonlinear Stochastic Differential Equations by a Local Linearization Method," *Stochastic Analysis and Applications*, 16, 733–752. [821]

Stramer, O., and Yan, J. (2006), "On Simulated Likelihood of Discretely Observed Diffusion Processes and Comparison to Closed-Form Approximation," technical report, University of Iowa. [822]

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. [832]