

Monte Carlo identities in the Path Space

Omiros Papaspiliopoulos

Universitat Pompeu Fabra

<http://www.econ.upf.edu/~omiros>

Joint work with **Gareth Roberts**, part of a SEMSTAT book
chapter

SAMSI, SMC 2008

Computational paradigm for simulation of stochastic processes

Instead of approximating the process (**discretizations**) and then design an MC scheme on the finite-dimensional discretization, design an MC scheme for the infinite-dimensional process and, if necessary, discretize the algorithm (**projection**). These operations typically do not commute.

Computational paradigm for simulation of stochastic processes

Instead of approximating the process (**discretizations**) and then design an MC scheme on the finite-dimensional discretization, design an MC scheme for the infinite-dimensional process and, if necessary, discretize the algorithm (**projection**). These operations typically do not commute.

Working on a discretization masks problems with the corresponding MC algorithms. Often **reducing the approximation bias** leads to **explosion of MC variance**. Additionally, **novel methodology** is generated by trying to design MC for the infinite-dimensional problem (e.g. non-centred parametrisations, retrospective sampling, MC with unbiased estimators of densities, hierarchical simulation models, coupling). Finally, the connection between **simulation** and **estimation** becomes transparent in that framework

To fix ideas we concentrate on a specific problem on simulation of **diffusion processes**. On its own has great interest (it is component of parameter and state estimation in continuous-time problems). Ideas however extend to other contexts

I will sketch main issues in MCMC and IS and then focus on the latter. Give some (very incomplete) indicative references

Discretely observed diffusion processes

Model described by the SDE

$$dX_s = b(X_s, s; \theta) ds + \sigma(X_s, s; \theta) dB_s,$$

observed at $X_0 = u, X_T = v$, with transition density for $s < t$,

$$p_{s,t}(u, v; \theta) = P[X_t \in dv \mid X_s = u; \theta] / dv$$

Drift and diffusion coefficient are allowed to depend on time (**inhomogeneous case**). However in my development for simplicity the main model will be **homogeneous**.

Transition density is **typically intractable**. A notable exception is **linear SDEs** where

$$b(x, t) = a(t)x + c(t)$$

$$\sigma(x, t) = \sigma(t)$$

Then $p_{s,t}(u, v; \theta)$ is a Gaussian density (still possibly not available in closed form)

Transition density is **typically intractable**. A notable exception is **linear SDEs** where

$$\begin{aligned}b(x, t) &= a(t)x + c(t) \\ \sigma(x, t) &= \sigma(t)\end{aligned}$$

Then $p_{s,t}(u, v; \theta)$ is a Gaussian density (still possibly not available in closed form)

Generally, approximations for small time increments are available, e.g Euler (Gaussian) approximation for homogeneous SDEs:

$$p_{0,\Delta}(u, v; \theta) \approx \phi(v; b(u; \theta)\Delta, \sigma(u; \theta)^2\Delta), \Delta \approx 0,$$

ϕ is the Gaussian density

Due to Markov structure I will only consider two consecutive points, X_0, X_T . Trivial to extend to arbitrary number

We are interested in dual problems (motivated by intractability of transition density for given T but tractability given high frequency data)

- ▶ Imputation: simulate the **missing data** $(X_s, s \in (0, T))$.
- ▶ Estimation: find MC approximations for the **likelihood** $p_{0,T}(u, v; \theta)$, or try to sample from the corresponding posterior distribution

We are interested in dual problems (motivated by intractability of transition density for given T but tractability given high frequency data)

- ▶ Imputation: simulate the **missing data** ($X_s, s \in (0, T)$).
- ▶ Estimation: find MC approximations for the **likelihood** $p_{0,T}(u, v; \theta)$, or try to sample from the corresponding posterior distribution

Discretization approach: work with the finite-dimensional vector $X^{(n)} = (X_0, X_\Delta, X_{2\Delta}, \dots, X_{n\Delta} = X_T)$ together with some approximation of the joint law, e.g. based on Euler approximation

$$p(X^{(n)} | X_0, \theta) = \prod_{i=1}^n \phi(X_{i\Delta}; b(X_{(i-1)\Delta}; \theta)\Delta, \sigma(X_{(i-1)\Delta}; \theta)^2 \Delta)$$

Discrete-time problem. Then take $\Delta \rightarrow 0$ (keeping $n\Delta = T$) to reduce approximation bias.

Typical problems with discretization approach

- ▶ MCMC1 (estimation): Suppose (for illustration) that

$$dX_s = b(X_s) ds + \theta dB_s$$

and we wish to sample from $p(\theta | X_0, X_T)$ using a **data augmentation**, i.e. iterative simulation of $X^{(n)}$ and θ . However, a quadratic variation identity shows that

$$\theta^2 = \frac{1}{T} \lim_{n \rightarrow \infty} \sum_i |X_{i\Delta} - X_{(i-1)\Delta}|^2$$

This implies that for large n the algorithm will never move. Similarly for an MCEM algorithm.

Typical problems with discretization approach

- ▶ MCMC1 (estimation): Suppose (for illustration) that

$$dX_s = b(X_s) ds + \theta dB_s$$

and we wish to sample from $p(\theta | X_0, X_T)$ using a **data augmentation**, i.e. iterative simulation of $X^{(n)}$ and θ . However, a quadratic variation identity shows that

$$\theta^2 = \frac{1}{T} \lim_{n \rightarrow \infty} \sum_i |X_{i\Delta} - X_{(i-1)\Delta}|^2$$

This implies that for large n the algorithm will never move. Similarly for an MCEM algorithm.

Roberts and Stramer (2001) provide a methodology based on factorization of diffusion measure. Part of a generic methodology for **decoupling** of stochastic processes from its parameters: **non-centred methodology** P, Roberts and Skold (2007)

- ▶ MCMC2 (imputation): Consider the simulation of $X^{(n)}$. This is also part of the previous algorithm. Necessary to do **block updates** due to dependence. A possibility is Metropolis-Hastings or Langevin algorithm. Understanding how to center the proposals so that the algorithm so that to have a valid algorithm in the $n \rightarrow \infty$ limit is non-trivial. See for example Stuart, Voss and Wiberg (2004)

- ▶ MCMC2 (imputation): Consider the simulation of $X^{(n)}$. This is also part of the previous algorithm. Necessary to do **block updates** due to dependence. A possibility is Metropolis-Hastings or Langevin algorithm. Understanding how to center the proposals so that the algorithm so that to have a valid algorithm in the $n \rightarrow \infty$ limit is non-trivial. See for example Stuart, Voss and Wiberg (2004)
- ▶ **Exact Simulation** (imputation): When considering the infinite-dimensional problem of simulating $(X_s, s \in (0, T))$ it turns out that in many families of diffusions it is possible to generate **exact skeletons** from the conditional distribution by (local) rejection sampling on the path space; Beskos, P and Roberts (2006,2008, current).

- ▶ MCMC2 (imputation): Consider the simulation of $X^{(n)}$. This is also part of the previous algorithm. Necessary to do **block updates** due to dependence. A possibility is Metropolis-Hastings or Langevin algorithm. Understanding how to center the proposals so that the algorithm so that to have a valid algorithm in the $n \rightarrow \infty$ limit is non-trivial. See for example Stuart, Voss and Wiberg (2004)
- ▶ **Exact Simulation** (imputation): When considering the infinite-dimensional problem of simulating $(X_s, s \in (0, T))$ it turns out that in many families of diffusions it is possible to generate **exact skeletons** from the conditional distribution by (local) rejection sampling on the path space; Beskos, P and Roberts (2006,2008, current).
(estimation): Also, examining the structure of the algorithm we've developed in Beskos, P, Roberts and Fearnhead (2006) a methodology for the **exact MCMC** for θ - **doubly intractable distributions**

- IS (estimation): **Simulated likelihood**. Given a model

$$dX_s = b(X_s; \theta) ds + \sigma(X_s; \theta) dB_s,$$

Using the Chapman-Kolmogorov and Euler approximation
Pedersen (1995) uses

$$\begin{aligned} p_{0,T}(u, v; \theta) &= E_X [p_{t,T}(X_t, v; \theta)] \\ &\approx E_X [\phi(v, b(X_t)(T-t), \sigma(X_t)^2(T-t))] \end{aligned}$$

This suggest an MC approximation where we generate
(unconditionally) paths up to time $t < T$ and weight them by

$$\phi(v, b(X_t)(T-t), \sigma(X_t)^2(T-t)) = \frac{1}{\sqrt{2\pi}\sigma(X_t)} \psi_t$$

where

$$\psi_t = \frac{1}{\sqrt{T-t}} \exp \left\{ -\frac{1}{2(T-t)\sigma(X_t)^2} (X_t - v)^2 \right\}$$

Clearly

$$E_X \left[\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \right] = 1, \forall t$$

but $\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \rightarrow 0$ a.s. The pathological behaviour of ψ_t is clear

Hence, trying to reduce the approximation error results in an explosion of the variance.

Clearly

$$E_X \left[\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \right] = 1, \forall t$$

but $\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \rightarrow 0$ a.s. The pathological behaviour of ψ_t is clear

Hence, trying to reduce the approximation error results in an explosion of the variance.

Additionally, in this approach it is not clear what is the connection between the methodology used to estimate the likelihood and the methodology for imputation of missing paths. We will see that approaches which operate on the path space resolve this issue and give estimates with finite variance in the limit.

Clearly

$$E_X \left[\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \right] = 1, \forall t$$

but $\frac{1}{\sqrt{2\pi\sigma(X_t)}} \psi_t \rightarrow 0$ a.s. The pathological behaviour of ψ_t is clear

Hence, trying to reduce the approximation error results in an explosion of the variance.

Additionally, in this approach it is not clear what is the connection between the methodology used to estimate the likelihood and the methodology for imputation of missing paths. We will see that approaches which operate on the path space resolve this issue and give estimates with finite variance in the limit.

Durham and Gallant (2002) in a stimulating and popular paper suggest an ad-hoc approach which gives good results. The continuous-time framework also clarifies why this is the case

IS and MC identities on the path space

We simplify the notation

$$dX_s = b(X_s) ds + \sigma(X_s) dB_s,$$

and to avoid notational complications we assume **scalar diffusions**. The general constructions extend directly to arbitrary finite dimensions but require technical conditions, which do not change the general picture. We also simplify the notation when $s = 0$

$$p_{0,t}(u, v; \theta) =: p_t(u, v)$$

IS and MC identities on the path space

We simplify the notation

$$dX_s = b(X_s) ds + \sigma(X_s) dB_s,$$

and to avoid notational complications we assume **scalar diffusions**. The general constructions extend directly to arbitrary finite dimensions but require technical conditions, which do not change the general picture. We also simplify the notation when $s = 0$

$$p_{0,t}(u, v; \theta) =: p_t(u, v)$$

Dual aim

1. IS approximation to the law of the **conditioned diffusion**, i.e. the process $(X_s, s \in (0, T))$ given $X_0 = u, X_T = v$.
2. MC estimators of $p_T(u, v)$ (simulated likelihood).

Background I: conditioned diffusions

The process X conditioned on its endpoints is still Markov, with a time-inhomogeneous SDE given by:

$$\begin{aligned}dX_s &= \tilde{b}(X_s, s) ds + \sigma(X_s) dB_s \\ \tilde{b}(x, t) &= b(x) + \sigma(x)^2 \nabla_x p_{t, T}(x, v)\end{aligned}$$

This is typically intractable, unless X is a **linear SDE**. Note the time-inhomogeneity

Background II: unconditional densities on the path space

Let $\mathbb{P}|_t$ be the measure of the process X on $[0, t)$, and $\mathbb{P}_0|_t$ the measure of the **driftless process**

$$dX_s = \sigma(X_s) dB_s,$$

Let $q_T(u, v)$ denote the transition density of the driftless process. Additionally let

$$A(x) = 1/\sigma(x)^2$$

Then, **Girsanov's theorem** gives the likelihood ratio

$$\begin{aligned} \frac{d\mathbb{P}}{d\mathbb{P}_0}|_t &= \exp \left\{ \int_0^t b(X_s) A(X_s) dX_s - \frac{1}{2} \int_0^t b(X_s)^2 A(X_s) ds \right\} \\ &=: \exp \{ G_t(X) \} \end{aligned}$$

Methodology based on conditioned processes

Let \mathbb{P}^* and \mathbb{P}_0^* be the measures of the conditioned processes. Consider the following heuristic argument given in Beskos et. al (2006) (which can be made rigorous, see Dacunha-Castelle & Florens-Zmirou (1986) and Deylon & Hu (2006))

By a **marginal-conditional** decomposition of measure we get

$$\frac{d\mathbb{P}}{d\mathbb{P}_0} = \frac{d\mathbb{P}^*}{d\mathbb{P}_0^*} \frac{p_T(u, v)}{q_T(u, v)}$$

Application of the decomposition

- ▶ IS for conditioned diffusions:

$$\frac{d\mathbb{P}^*}{d\mathbb{P}_0^*}(X) = \exp\{G_T(X)\} \frac{q_T(u, v)}{p_T(u, v)} \propto \exp\{G_T(X)\}$$

Application of the decomposition

- ▶ IS for conditioned diffusions:

$$\frac{d\mathbb{P}^*}{d\mathbb{P}_0^*}(X) = \exp\{G_T(X)\} \frac{q_T(u, v)}{p_T(u, v)} \propto \exp\{G_T(X)\}$$

- ▶ Identity for the transition density:

$$p_T(u, v) = q_T(u, v) E_{\mathbb{P}_0^*} [\exp\{G_T(X)\}]$$

These expressions are useful when

1. it is easy to simulate from \mathbb{P}_0^*
2. $q_T(u, v)$ is tractable

The only processes for which 1 and 2 are satisfied is when \mathbb{P}_0 is the measure of a **linear SDE**. Linear SDEs correspond to **Gaussian processes** hence the law of the conditioned process is a tractable Gaussian (and Markov) process. **Brownian bridge**

On the basis of the original model (11) we need that the diffusion coefficient is state-independent. Most one-dimensional diffusions can be transformed to have this property. In higher dimensions the approach is more restrictive

Assuming for simplicity that $\sigma = 1$, the methodology based on conditioned processes gives us as proposal the Brownian bridge

$$dX_t = \frac{X_t - v}{T - t} dt + dB_t$$

- ▶ IS approximation to the law of \mathbb{P}^* by path particles $(X^{(i)}, w(X^{(i)}))_{i=1}^N$, where X is drawn from the Gaussian measure \mathbb{P}_0^* , and

$$w(X^{(i)}) = \frac{\exp\{G_T(X^{(i)})\}}{\sum_{i=1}^N \exp\{G_T(X^{(i)})\}}$$

Assuming for simplicity that $\sigma = 1$, the methodology based on conditioned processes gives us as proposal the Brownian bridge

$$dX_t = \frac{X_t - v}{T - t} dt + dB_t$$

- ▶ IS approximation to the law of \mathbb{P}^* by path particles $(X^{(i)}, w(X^{(i)}))_{i=1}^N$, where X is drawn from the Gaussian measure \mathbb{P}_0^* , and

$$w(X^{(i)}) = \frac{\exp\{G_T(X^{(i)})\}}{\sum_{i=1}^N \exp\{G_T(X^{(i)})\}}$$

- ▶ MC **unbiased** estimator for the transition density

$$p_T(u, v) = q_T(u, v) \sum_{i=1}^N \exp\{G_T(X^{(i)})\}$$

The last expression is first derived in Dacunha-Castelle & Florens-Zmirou (1986). It has then be re-discovered and even re-proved various times, most without even understanding the connection. It is also easy to get unbiased estimators of the score function

Practical implementation

There are various routes which one can take trying to use mathematical formulation in (17)

Practical application might require at this stage **approximation** of the weights, e.g. based on a Riemann approximation, and simulation of a finite-dimensional skeleton of X , which in this case can be done exactly.

Practical implementation

There are various routes which one can take trying to use mathematical formulation in (17)

Practical application might require at this stage **approximation** of the weights, e.g. based on a Riemann approximation, and simulation of a finite-dimensional skeleton of X , which in this case can be done exactly.

However other more interesting possibilities exist

Exact inference

A further product of this development is discovered in Beskos et. al (2005,2008) and Fearnhead et.al (2008). Under further structure on the drift $b(x)$ (which trivially incorporates one-dimensional diffusions) it is shown how to

1. Simulate exactly from \mathbb{P}^* based on **retrospective rejection sampling** based on the derived likelihood ratio

Exact inference

A further product of this development is discovered in Beskos et. al (2005,2008) and Fearnhead et.al (2008). Under further structure on the drift $b(x)$ (which trivially incorporates one-dimensional diffusions) it is shown how to

1. Simulate exactly from \mathbb{P}^* based on **retrospective rejection sampling** based on the derived likelihood ratio
2. Generate **unbiased estimators** of $w(X)$ rather than approximating. Random weight IS.
 - ▶ It is based on a generic methodology which gets UE based on series expansions.
 - ▶ This approach corresponds to an **exact discrete-time representation** of the process. A methodology for **MC using unbiased estimators of densities**. Another instance of the principle in Arnaud's talk
 - ▶ This methodology can be directly feed into an SMC framework, time permitting I will show the main construction

Methodology based on guided processes

This methodology is applied when the conditioned driftless process is intractable. Hence it is particularly relevant in multivariate problems.

Some details are developed in Delyon & Hu (2006), a heuristic approach was first developed in Durham & Gallant (2002)

The idea is to find the LR of \mathbb{P}^* wrt to a process explicitly constructed to hit the end-point, rather than probabilistically conditioned to do so. Hence we call them **guided processes**

Methodology based on guided processes

This methodology is applied when the conditioned driftless process is intractable. Hence it is particularly relevant in multivariate problems.

Some details are developed in Delyon & Hu (2006), a heuristic approach was first developed in Durham & Gallant (2002)

The idea is to find the LR of \mathbb{P}^* wrt to a process explicitly constructed to hit the end-point, rather than probabilistically conditioned to do so. Hence we call them **guided processes**

Two such options

$$(DG) \quad dY_t = \frac{Y_t - v}{T - t} dt + \sigma(Y_t) dB_t, \quad \mathbb{Z}_t$$

$$(DH) \quad dY_t = \frac{Y_t - v}{T - t} dt + b(Y_t) dt + \sigma(Y_t) dB_t, \quad \mathbb{L}_t$$

Sketch of the technique

Extension of regular Girsanov allows us to define for $t < T$

$$\frac{d\mathbb{P}^*}{d\mathbb{Z}} \Big|_t \text{ or } \frac{d\mathbb{P}^*}{d\mathbb{L}} \Big|_t$$

then take the limit as $t \rightarrow T$ and using dominated convergence show that the limiting weights define the change of measure

Key to the technique is an integration by parts which expresses the weight as a function of ψ_t (9) the leading order term which makes Pedersen's method collapse

IS and MC identities

Let

$$\varphi_t = \exp \left\{ - \int_0^t \frac{(Y_s - v)^2 A(Y_s)}{T - s} b(Y_s) ds - \frac{1}{2} \int_0^t \frac{(Y_s - v)^2}{T - s} \diamond dA(Y_s) \right\}$$

$$\tilde{\varphi}_t = \exp \left\{ - \frac{1}{2} \int_0^t \frac{(Y_s - v)^2}{T - s} \diamond dA(Y_s) + G_t(Y) \right\}$$

IS and MC identities

Let

$$\varphi_t = \exp \left\{ - \int_0^t \frac{(Y_s - v)^2 A(Y_s)}{T - s} b(Y_s) ds - \frac{1}{2} \int_0^t \frac{(Y_s - v)^2}{T - s} \diamond dA(Y_s) \right\}$$

$$\tilde{\varphi}_t = \exp \left\{ - \frac{1}{2} \int_0^t \frac{(Y_s - v)^2}{T - s} \diamond dA(Y_s) + G_t(Y) \right\}$$

Then, main result for an arbitrary function f we have the change of measure

$$E_X[f(X) \mid X_T = v] = \frac{E_Y[f(Y)\varphi_T]}{E_Y[\varphi_T]} = \frac{E_Y[f(Y)\tilde{\varphi}_T]}{E_Y[\tilde{\varphi}_T]}$$

and crucially

$$p_T(u, v) = \phi(v; u, T\sigma(u)^2) \frac{\sigma(v)}{\sigma(u)} E_Y[\tilde{\varphi}_T]$$

Note the simplification to the conditioned process case when σ is constant.

For practical purposes one might approximate now the weights AND the process. There is no unbiased estimation anymore

Note that the guided process is not typically conditioned version of another process...

One can also extend these ideas for defining MCMC on the path space. This is further explored on the paper. MCMC approaches along these lines (based on discretizations) are developed for example in Golightly and Wilkinson (2008), Chib, Pitt and Shephard (2007).

Random weight PF for partially observed diffusions

For a wide class of diffusion models we can use the methodology based on IS using conditioned processes to yield novel PF algorithms for partially observed diffusions models.

Further simplify the notation, and state the problem in its most basic form: **Observations** y_i at time t_i , which relate with an **unobserved diffusion** $X_{t_i} =: x_i$ via a likelihood function $p(y_i | x_i)$.

Let $p(x_i | x_{i-1})$ be the transition density

Aim is to estimate sequentially and in real-time the state of the unobserved process X at the observation times, given all available data up to that time. **Filtering distributions**

$$\pi_{i+1}(x_{i+1}) := p(x_{i+1} | y_{1:i+1}), \quad y_{1:i+1} = (y_1, \dots, y_{i+1})$$

Auxiliary Particle Filter

Let $\{(x_i^{(j)}, w_i^{(j)})\}_{j=1}^N$ be a particle (importance sampling) approximation of π_i .

Choose a proposal on the joint discrete-continuous space:

$$g(x, k) \propto \beta_k q(x | x_i^{(k)}, y_{i+1}), x \in R^d, k = 1 : N$$

where q and β are user-chosen proposals (which incorporate the information of y_{i+1}).

Then, the weight corresponding to a new particle x_{i+1} simulated according to this proposal is calculated as

$$w_{i+1} = \frac{w_i^{(k)} p(y_{i+1} | x_{i+1}) p(x_{i+1} | x_i^{(k)})}{\beta_k q(x_{i+1} | x_i^{(k)}, y_{i+1})}$$

Then $\{(x_{i+1}^{(j)}, w_{i+1}^{(j)})\}_{j=1}^N$ is a particle approximation of π_{i+1} .

Particle filtering for diffusions: major challenge

The weight associated with particle x_{i+1} is **intractable**

$$\frac{w_i^{(k)} p(y_{i+1}|x_{i+1}) p(x_{i+1}|x_i^{(k)})}{\beta_k q(x_{i+1}|x_i^{(k)}, y_{i+1})}$$

Our solution: Random weight particle filter I

We replace the intractable $p(x_{i+1}|x_i)$ by a **positive unbiased estimator**. “Properly weighted IS”

We construct some **auxiliary variables** V_{i+1} from an appropriate distribution $Q_{i+1}(V_{i+1} | x_i, x_{i+1})$, and a **positive** function $r(V_{i+1}, x_i, x_{i+1}, t_i, t_{i+1}) \geq 0$, s.t.

$$E[r(V_{i+1}, x_i, x_{i+1})] = \int r(u, x_i, x_{i+1}) dQ_{i+1}(u | x_i, x_{i+1}) = p(x_{i+1}|x_i)$$

Easy to simulate from Q_{i+1} and to evaluate r .

Random weight particle filter II

Simple modification of the auxiliary particle filter:

1. Simulate $(x_{i+1}, x_i^{(k)})$ from $\beta_k q(x_{i+1} | x_i^{(k)}, y_{i+1})$
2. Simulate v_{i+1} from $Q_{i+1}(V_{i+1} | x_i^{(k)}, x_{i+1}, t_i, t_{i+1})$
3. Compute the weight

$$w_{i+1} = \frac{w_i^{(k)} p(y_{i+1} | x_{i+1}) r(v_{i+1}, x_i^{(k)}, x_{i+1})}{\beta_k q(x_{i+1} | x_i^{(k)}, y_{i+1})} \quad (1)$$

Then, it is easy to show that $\{(x_{i+1}^{(j)}, w_{i+1}^{(j)})\}_{j=1}^N$ is a particle approximation of $\tilde{\pi}_{i+1}$.

MC with unbiased estimates of densities

The idea of replacing in standard MC methods (e.g. IS, MCMC) intractable densities with unbiased estimators is intriguing.

In a disguised form it is already in earlier PF implementations (e.g. auxiliary PF, partial rejection control).

In Beskos et al (2006) we exploited such constructions for exact MCMC inference for diffusions

Independently this idea has been developed recently in various fields facing this intractability, e.g. ABC, Expected Auxiliary method. Particularly interesting direction: Particle MCMC.

There is a nice intuition on the effect of adding the auxiliary simulations into the calculation of the IS weights

Discrete-time representation of the state

- ▶ Conditionally on X_{t_i} , $X_{t_{i+1}}$, t_i and t_{i+1} , V_{i+1} is independent of V_j and X_{t_j} for any j different from $i, i + 1$.
- ▶ It follows from the unbiasedness and positivity of r that, conditionally on $X_{t_i} = x$, $r(v_{i+1}, x, x_{t_{i+1}}, t_i, t_{i+1})$ is a probability density function for $(X_{t_{i+1}}, V_{i+1})$ with respect to the product measure $Leb(dz) \times Q_{i+1}(dv \mid x, z)$, where Leb denotes the Lebesgue measure.

- ▶ Hence, we get an equivalent discrete-time representation of the discretely-sampled diffusion, as **marginal state** of a higher-dimensional discrete-time model with states $(Z_i, V_i), i = 1, \dots, n, Z_i \in \mathbf{R}^d$, with a non-homogeneous **tractable** transition density between consecutive states

$$p_{i+1}(z_{i+1}, v_{i+1} \mid z_i, v_i) = r(v_{i+1}, z_i, z_{i+1}, t_i, t_{i+1}),$$

Hierarchical models representation; variance considerations

- ▶ Hence, we get an equivalent discrete-time representation of the discretely-sampled diffusion, as **marginal state** of a higher-dimensional discrete-time model with states $(Z_i, V_i), i = 1, \dots, n, Z_i \in \mathbf{R}^d$, with a non-homogeneous **tractable** transition density between consecutive states

$$p_{i+1}(z_{i+1}, v_{i+1} \mid z_i, v_i) = r(v_{i+1}, z_i, z_{i+1}, t_i, t_{i+1}),$$

Hierarchical models representation; variance considerations

- ▶ In the filtering problem, we get data y_i with observation density $p(y_{i+1} \mid z_{i+1})$, hence **discrete-discrete** filtering problem.
- ▶ Consider an auxiliary particle filter generating new particles according to the following proposal probability measure

$$(z_{i+1}, v_{i+1}) \sim q(z_{i+1} \mid z_i^{(k)}, y_{i+1}) Q_{i+1}(dv_{i+1}) \text{Leb}(dz_{i+1}),$$

The weights associated with each particle in this discrete-time model are tractable and are given by (1).