Numerics for McKean-Vlasov SDEs with superlinear growth

Gonçalo dos Reis
University of Edinburgh (UK) & CMA/FCT/UNL (PT)

S. Engelhardt (Jena) & G. Smith (Moody's Analytics) & P. Tankov (ENSAE)

Nonlinear Processes and their Applications

St. Etienne, 05 Jul 2019

Partial funding by UID/MAT/00297/2019

Outline

Outline

McKean-Vlasov stochastic differential equations

MV-SDE are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 = x, \qquad (MV - SDE)$$

where μ_t is the law of X_t, and W is a standard BM.

McKean-Vlasov stochastic differential equations

MV-SDE are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 = x, \qquad (MV - SDE)$$

where μ_t is the law of X_t, and W is a standard BM.

Example (Mean Field Scalar Interaction)

Let us consider a simple example:

$$X(t) = X_0 + \int_0^t \Big[\mathbb{E}[X(s)] + X(s) - X^3(s)\Big] ds + W(t)$$

McKean-Vlasov stochastic differential equations

MV-SDE are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 = x, \qquad (MV - SDE)$$

where μ_t is the law of X_t, and W is a standard BM.

Example (Mean Field Scalar Interaction)

Let us consider a simple example:

$$X(t) = X_0 + \int_0^t \Big[\mathbb{E}[X(s)] + X(s) - X^3(s)\Big] ds + W(t)$$

Assumption: We assume time-continuity and the following:

1. σ is Lipschitz in x and in law for the Wasserstein distance $W^{(2)}(\mu, \mu')$.
2. b is Lipschitz in law and has monotone growth in x:

$$\langle x - x', b(t, x, \mu) - b(t, x', \mu)\rangle \leq L|x - x'|^2.$$

3. b is locally Lipschitz with polynomial growth in x: $\exists q > 1 : \forall t \in [0, T]$:

$$|b(t, x, \mu) - b(t, x', \mu)| \leq L(1 + |x|^q + |x'|^q)|x - x'|.$$

Under these assumptions, (MV-SDE) has a unique strong solution.

Approximation of MV-SDE

A common technique for simulating MV-SDEs: interacting particle system:

$$\mathrm{d}X_t^{i,N} = b\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i,$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \ldots, N$ are independent.

Approximation of MV-SDE

A common technique for simulating MV-SDEs: interacting particle system:

$$\mathrm{d}X_t^{i,N} = b\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i,$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \dots, N$ are independent.

"Propagation of chaos" (Sznitman '91) : under appropriate conditions, as $N \to \infty$, for every i, the process $X^{i,N}$ converges to X^i, the solution of the MV-SDE driven by the Brownian motion W^i.

$$\lim_{N\to\infty} \sup_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^{i,N} - X_t^i|^2\Big] = 0\,.$$

This system of ordinary SDEs can be discretized with, e.g., the Euler scheme. Under global Lipschitz conditions, one can then prove the following convergence result (e.g. Carmona'16)

Approximation of MV-SDE

We have the following result.

Theorem (dRES2018, Propagation of chaos)

Let $X_0 \in L_0^m(\mathbb{R}^d)$ hold for $m > 4$. Then for superlinear growth MV-SDEs

$$\sup_{1\leq i\leq N} \mathbb{E}[\sup_{0\leq t\leq T} |X_t^i - X_t^{i,N}|^2] \leq C \begin{cases} N^{-1/2} & \text{if } d < 4, \\ N^{-1/2}\log(N) & \text{if } d = 4, \\ N^{-2/d} & \text{if } d > 4. \end{cases}.$$

Therefore, to show convergence between our numerical scheme and the MV-SDE, we only need to show that the "true" particle scheme and numerical version of the particle scheme converge.

Approximation of MV-SDE

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

$$\hat{\theta}^{N,n} = \frac{1}{N} \sum_{i=1}^{N} G(X^{i,N,n}).$$

Approximation of MV-SDE

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

$$\hat{\theta}^{N,n} = \frac{1}{N}\sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.

Approximation of MV-SDE

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

$$\hat{\theta}^{N,n} = \frac{1}{N}\sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.
- The discretization error: difference between $\mathbb{E}[G(X^{i,N,n})]$ and $\mathbb{E}[G(X^{i,N})]$. Under Lipschitz assumptions the Euler scheme has weak error of order $\frac{1}{n}$.

Approximation of MV-SDE

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

$$\hat{\theta}^{N,n} = \frac{1}{N} \sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.
- The discretization error: difference between $\mathbb{E}[G(X^{i,N,n})]$ and $\mathbb{E}[G(X^{i,N})]$. Under Lipschitz assumptions the Euler scheme has weak error of order $\frac{1}{n}$.
- The propagation of chaos error: difference between $\mathbb{E}[G(X^{i,N})]$ and $\mathbb{E}[G(X)]$. For G and X nice enough this error is also of order $\frac{1}{\sqrt{N}}$

  See Kohatsu-Higa and Ogawa, '97, Bossy '04.

Our questions

MV-SDE are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 = x, \qquad (MV - SDE)$$

where $\mu_t = \mathbb{P} \circ X_t^{(-1)}$ under super-linear growth of the drift.

Our questions

MV-SDE are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 = x, \qquad (MV-SDE)$$

where $\mu_t = \mathbb{P} \circ X_t^{(-1)}$ under super-linear growth of the drift.

▷ **Our research questions:**

- Euler scheme with *standard* convergence rate ?
  - Many results for Lipschitz case
- Importance sampling for some $\mathbb{E}[G(X)]$?
  - Empty for IS. Existing results for other variance reduction techniques.

Outline

MV-SDEs wit super linear growth and standard Euler

The MV-SDE for $p \geq 2$

$$\mathrm{d}X_t = b(t, X_t, \mu_t^X)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

The particle approximation

$$\mathrm{d}X_t^{i,N} = b\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i,$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \dots, N$ are independent.

MV-SDEs wit super linear growth and standard Euler

The MV-SDE for $p \geq 2$

$$\mathrm{d}X_t = b(t, X_t, \mu_t^X)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

The particle approximation

$$\mathrm{d}X_t^{i,N} = b\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i,$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \ldots, N$ are independent.
And the **explicit** Euler scheme:

$$\bar{X}_{t_{k+1}}^{i,N,M} = \bar{X}_{t_k}^{i,N,M} + b\Big(t_k, \bar{X}_{t_k}^{i,N,M}, \bar{\mu}_{t_k}^{X,N}\Big)h + \sigma\Big(t_k, \bar{X}_{t_k}^{i,N,M}, \bar{\mu}_{t_k}^{X,N}\Big)\Delta W_{t_k}^i,$$

where $\bar{\mu}_{t_k}^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\bar{X}_{t_k}^{j,N,M}}(\mathrm{d}x)$, $\Delta W_{t_k}^i := W_{t_{k+1}}^i - W_{t_k}^i$ and $\bar{X}_0^{i,N,M} := X_0^i$ (random initial condition).

Euler goes wrong

The stochastic Ginzburg Landau equation and with added mean field term,

$$\mathrm{d}X_t = \Big(\frac{\sigma^2}{2}X_t - X_t^3 + c\mathbb{E}[X_t]\Big)\mathrm{d}t + \sigma X_t \mathrm{d}W_t, \quad X_0 = x.$$

Euler goes wrong

The stochastic Ginzburg Landau equation and with added mean field term,

$$\mathrm{d}X_t = \Big(\frac{\sigma^2}{2}X_t - X_t^3 + c\mathbb{E}[X_t]\Big)\mathrm{d}t + \sigma X_t \mathrm{d}W_t, \quad X_0 = x.$$

We simulate $N = 5000$ particles with a time step $h = 0.05$, $T = 2$ and $X_0 = 1$, we also take $\sigma = 3/2$ and $c = 1/2$.

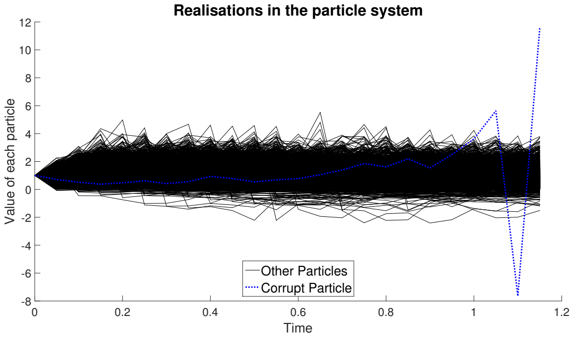

Figure: The dashed particle is starting to oscillate and is taking larger values than it surrounding particles.

Euler goes wrong

- All particles are reasonably well behaved until one starts to oscillate rapidly, then the whole system diverges.
- This happens for the Euler scheme via a handful of Monte Carlo simulations that return extremely large (or infinite) values. These few events dominate the others: an *exponentially small probability event has a double exponential impact*.

Euler goes wrong

- All particles are reasonably well behaved until one starts to oscillate rapidly, then the whole system diverges.
- This happens for the Euler scheme via a handful of Monte Carlo simulations that return extremely large (or infinite) values. These few events dominate the others: an *exponentially small probability event has a double exponential impact*.
- More particles could reduce the dependency between them and hence influence the system less.

Euler goes wrong

- All particles are reasonably well behaved until one starts to oscillate rapidly, then the whole system diverges.
- This happens for the Euler scheme via a handful of Monte Carlo simulations that return extremely large (or infinite) values. These few events dominate the others: an *exponentially small probability event has a double exponential impact*.
- More particles could reduce the dependency between them and hence influence the system less.
  Its the opposite!.
  Same example but with $N = [1000, 5000, 10000, 20000]$ particles and rerun each case 1000 times and record the total number of times we observe a divergence over the ensemble.

Number of particles	1000	5000	10000	20000
Number of blow ups	3	32	43	108

Table: Number of divergences recorded at each particle level out of 1000 simulations.

Direct Euler doesn't work: try implicit scheme

We also consider the following implicit scheme

$$\tilde{X}^{i,N,M}_{t_{i+1}} = \tilde{X}^{i,N,M}_{t_i} + b\Big(t_i, \tilde{X}^{i,N,M}_{t_{i+1}}, \tilde{\mu}^{X,N,M}_{t_i}\Big)h + \sigma\Big(t_i, \tilde{X}^{i,N,M}_{t_i}, \tilde{\mu}^{X,N,M}_{t_i}\Big)\Delta W^i_{t_i}, \qquad (1)$$

where $\tilde{\mu}^{X,N,M}_{t_k}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{\tilde{X}^{j,N,M}_{t_k}}(\mathrm{d}x)$ and $\tilde{X}^{i,N,M}_0 = X^i_0$.

- implicit equation to be solved at each step.
  Not implicit in the measure component.

Convergence of the implicit scheme

Theorem (Convergence of the Time-discretization)

Assume: Monotonicity+Holder in Time+$X_0 \in L^{4(q+1)}(\mathbb{R}^d)$*, and exists* $C > 0$ *for all* $\mu \in \mathcal{P}_2(\mathbb{R}^d)$*,*

$$|b(0,0,\mu)| + |\sigma(0,0,\mu)| \leq C\,.$$

Fix timestep $h^* < 1/\max(L_b, 2\beta)$*, then, for any* $T = Mh$ *and* $s \in [1,2)$

$$\sup_{1\leq i\leq N} \lim_{h\to 0} \mathbb{E}[|X_T^{i,N} - \tilde{X}_T^{i,N,M}|^s] = 0\,.$$

Theorem (Strong Convergence of Implicit Scheme)

Under the above conditions and σ *is only a function of time and space (does not have a measure dependence), for any* $T = Mh$ *and* $s \in [1,2)$ *one has*

$$\lim_{N\to\infty} \sup_{1\leq i\leq N} \lim_{h\to 0} \mathbb{E}[|X_T^{i} - \tilde{X}_T^{i,N,M}|^s] = 0\,.$$

Tamed Euler scheme

Inspired by the "taming" literature (hutzentaler, Jentzen, Sabanis, etc...) we consider a so-called *tamed* Euler explicit scheme.

With the notation above consider the following scheme

$$\bar{X}^{i,N,M}_{t_{k+1}} = \bar{X}^{i,N,M}_{t_k} + \frac{b\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)}{1 + M^{-\alpha}\Big|b\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)\Big|} h + \sigma\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)\Delta W^i_{t_k},$$

where $\bar{\mu}^{X,N}_{t_k}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}^{j,N,M}_{t_k}}(\mathrm{d}x)$ and $\alpha \in (0, 1/2]$ with $\bar{X}^{i,N,M}_0 = X^i_0$.

Convergence results

This then leads to our main explicit scheme convergence result,no boundedness assumptions.

Theorem (Strong Convergence of Explicit)

Under monotonicity + Holder in time hold + $X_0 \in L^m(\mathbb{R}^d)$ *for* $m \geq 4(1+q)$ *(note* $q > 1$*).*
Let X^i *be the solution to the MV-SDE (driven by* W^i*), and* $X^{i,N,M}$ *be the tamed schemes. Then we obtain the following convergence result*

$$\sup_{1\leq i\leq N} \mathbb{E}\Big[\sup_{0\leq t\leq T} |X_t^i - X_t^{i,N,M}|^2\Big] \leq C \begin{cases} h + N^{-1/2} & \text{if } d < 4, \\ h + N^{-1/2}\log(N) & \text{if } d = 4, \\ h + N^{-2/d} & \text{if } d > 4. \end{cases}$$

We unavoidably show the Monte Carlo time-discretization convergence rate

$$\sup_{1\leq i\leq N} \mathbb{E}[\sup_{0\leq t\leq T} |X_t^{i,N} - X_t^{i,N,M}|^2] \leq Ch.$$

We consider the Ginzburg-Landau eq with mean term (with $T = 1$). We then consider a set of dimensions from 1 to 200 and number of particles from 100 to 20000.

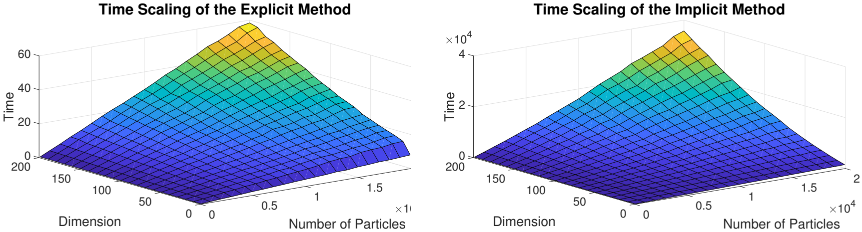

Figure: Showing how the time (in seconds) of the explicit scheme (left; timescale ≈ 60 seconds) and implicit scheme (right; timescale $\approx 10^4$ seconds) changes with particles and dimension.

Outline

Importance sampling

We focus on importance sampling, a technique to reduce statistical error based on the identity:

$$\mathbb{E}\Big[G(X)\Big] = \mathbb{E}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big].$$

Importance sampling

We focus on importance sampling, a technique to reduce statistical error based on the identity:

$$\mathbb{E}\Big[G(X)\Big] = \mathbb{E}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big].$$

Goal: Find importance sampling measure $\mathbb{Q}$ under which X and $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$ are easy to simulate and corresponding estimator has a smaller variance:

$$\textbf{Goal:} \quad \text{minimize} \quad \mathbb{E}\left[G^2(X)\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right].$$

Importance sampling

We focus on importance sampling, a technique to reduce statistical error based on the identity:

$$\mathbb{E}\Big[G(X)\Big] = \mathbb{E}\Big[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big] = \mathbb{E}_{\mathbb{Q}}\Big[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\Big].$$

Goal: Find importance sampling measure $\mathbb{Q}$ under which X and $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$ are easy to simulate and corresponding estimator has a smaller variance:

$$\textbf{Goal:} \quad \text{minimize} \quad \mathbb{E}\left[G^2(X)\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right].$$

This is a *generally intractable* problem and in the SDE framework one focuses on Cameron-Martin transforms:

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_T} = \mathcal{E}\Big(\int_0^T f_t \mathrm{d}W_t\Big) := \exp\left(\int_0^T f_t \mathrm{d}W_t - \frac{1}{2}\int_0^T f_t^2 \mathrm{d}t\right),$$

where f_t is a deterministic $\frac{1}{2}$-Hölder function with square integrable density.

Outline

The decoupling algorithm

Recall our goal: evaluate $\mathbb{E}[G(X)]$, where X is the solution to MV-SDE.

1. Use an N-particle system to approximate the empirical law μ_t^N (under $\mathbb{P}$). Define a new SDE, approximating the original MV-SDE:

$$\mathrm{d}\bar{X}_t = b(t, \bar{X}_t, \mu_t^N)\mathrm{d}t + \sigma(t, \bar{X}_t, \mu_t^N)\mathrm{d}W_t^{\mathbb{P}}, \quad \bar{X}_0 = x_0,$$

where $W^{\mathbb{P}}$ is a $\mathbb{P}$-BM independent of the BMs driving the particle system.

The decoupling algorithm

Recall our goal: evaluate $\mathbb{E}[G(X)]$, where X is the solution to MV-SDE.

1. Use an N-particle system to approximate the empirical law μ_t^N (under $\mathbb{P}$). Define a new SDE, approximating the original MV-SDE:

$$\mathrm{d}\bar{X}_t = b(t, \bar{X}_t, \mu_t^N)\mathrm{d}t + \sigma(t, \bar{X}_t, \mu_t^N)\mathrm{d}W_t^{\mathbb{P}}, \quad \bar{X}_0 = x_0,$$

   where $W^{\mathbb{P}}$ is a $\mathbb{P}$-BM independent of the BMs driving the particle system.

2. Change the probability measure to importance sampling measure $\mathbb{Q}$.

$$\mathrm{d}\bar{X}_t = \Big(b(t, \bar{X}_t, \mu_t^N) + \dot{h}_t\sigma(t, \bar{X}_t, \mu_t^N)\Big)\,\mathrm{d}t + \sigma(t, \bar{X}_t, \mu_t^N)\mathrm{d}W_t^{\mathbb{Q}}, \quad \bar{X}_0 = x_0\,.$$

The decoupling algorithm

Recall our goal: evaluate $\mathbb{E}[G(X)]$, where X is the solution to MV-SDE.

1. Use an N-particle system to approximate the empirical law μ_t^N (under $\mathbb{P}$). Define a new SDE, approximating the original MV-SDE:

$$\mathrm{d}\bar{X}_t = b(t, \bar{X}_t, \mu_t^N)\mathrm{d}t + \sigma(t, \bar{X}_t, \mu_t^N)\mathrm{d}W_t^{\mathbb{P}}, \quad \bar{X}_0 = x_0,$$

   where $W^{\mathbb{P}}$ is a $\mathbb{P}$-BM independent of the BMs driving the particle system.

2. Change the probability measure to importance sampling measure $\mathbb{Q}$.

$$\mathrm{d}\bar{X}_t = \Big(b(t, \bar{X}_t, \mu_t^N) + \dot{h}_t\sigma(t, \bar{X}_t, \mu_t^N)\Big)\,\mathrm{d}t + \sigma(t, \bar{X}_t, \mu_t^N)\mathrm{d}W_t^{\mathbb{Q}}, \quad \bar{X}_0 = x_0\,.$$

3. Compute $\mathbb{E}[G(X)] = \mathbb{E}_{\mathbb{Q}}\left[\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}G(X)\right]$ by Monte Carlo .

Implicit assumption: coefficients are easier to approximate than $\mathbb{E}[G(X)]$.

The complete measure change

1. Change the measure in the MV-SDE

$$dX_t = (b(t, X_t, \mu_t^{\mathbb{P}}) + \sigma(t, X_t, \mu_t^{\mathbb{P}})\dot{h}_t)dt + \sigma(t, X_t, \mu_t^{\mathbb{P}})\mathrm{d}W_t^{\mathbb{Q}}$$
$$dZ_t = \dot{h}_t Z_t \mathrm{d}W_t^{\mathbb{Q}}, \qquad \mu_t^{\mathbb{P}}(A) = \mathbb{E}^{\mathbb{Q}}[Z_t \mathbf{1}_A(X_t)].$$

The complete measure change

1. Change the measure in the MV-SDE

$$dX_t = (b(t, X_t, \mu_t^{\mathbb{P}}) + \sigma(t, X_t, \mu_t^{\mathbb{P}})\dot{h}_t)dt + \sigma(t, X_t, \mu_t^{\mathbb{P}})\mathrm{d}W_t^{\mathbb{Q}}$$
$$dZ_t = \dot{h}_t Z_t \mathrm{d}W_t^{\mathbb{Q}}, \qquad \mu_t^{\mathbb{P}}(A) = \mathbb{E}^{\mathbb{Q}}[Z_t \mathbf{1}_A(X_t)].$$

2. Approximate with particle system

$$\mathrm{d}X_t^{i,N} = \left(b\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N} Z_t^j \delta_{X_t^{j,N}}\right) + \dot{h}_t \sigma\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N} Z_t^j \delta_{X_t^{j,N}}\right)\right)\mathrm{d}t$$
$$+ \sigma\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N} Z_t^j \delta_{X_t^{j,N}}\right)\mathrm{d}W_t^{i,\mathbb{Q}},$$

$$\mathrm{d}Z_t^i = \dot{h}_t Z_t^i \mathrm{d}W_t^{i,\mathbb{Q}}, \quad Z_0^i = 1 \qquad \longrightarrow \boxed{\hat{\theta}_h = \frac{1}{N}\sum_{i=1}^{N} Z_T^{i,N} G(X_T^{i,N}).}$$

Implicit assumption: $\mathbb{Q}$ does not worsen estimation of coefficients.
(One can show Chaos propagation for the system under $\mathbb{Q}$.)

Outline

Minimizing the variance

- Optimality of measure changes come from LDPs + Varadhan's lemma.
  - ▷ Requires LDP for underlying process (in path space), here the BM

Minimizing the variance

- Optimality of measure changes come from LDPs + Varadhan's lemma.
  - Requires LDP for underlying process (in path space), here the BM
- To ensure pathwise continuity of solution with respect to the driving BM (to use contraction principle for LDPs), consider MV-SDE with constant volatility

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0,$$

where b satisfies monotone growth and local Lipschitz conditions.

Minimizing the variance

- Optimality of measure changes come from LDPs + Varadhan's lemma.
  - ▷ Requires LDP for underlying process (in path space), here the BM
- To ensure pathwise continuity of solution with respect to the driving BM (to use contraction principle for LDPs), consider MV-SDE with constant volatility
$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0,$$
where b satisfies monotone growth and local Lipschitz conditions.
- Pay-off functional G non-negative, continuous and satisfies growth condition
$$\log\big(G(x)\big) \leq C_1 + C_2 \sup_{t\in[0,T]} |x_t|^\alpha, \quad \alpha < 2,$$

Minimizing the variance

- Optimality of measure changes come from LDPs + Varadhan's lemma.
  - ▷ Requires LDP for underlying process (in path space), here the BM
- To ensure pathwise continuity of solution with respect to the driving BM (to use contraction principle for LDPs), consider MV-SDE with constant volatility

$$\mathrm{d}X_t = b(t, X_t, \mu_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0,$$

  where b satisfies monotone growth and local Lipschitz conditions.
- Pay-off functional G non-negative, continuous and satisfies growth condition

$$\log\big(G(x)\big) \leq C_1 + C_2 \sup_{t\in[0,T]} |x_t|^{\alpha}, \quad \alpha < 2,$$

- Denote by $\overline{G}$ the pay-off as functional of Brownian trajectory.

Optimal IS: the decoupling algorithm

We compute the optimal importance sampling measure for SDE

$$\mathrm{d}\overline{X}_t = b(t, \overline{X}_t, \mu_t^N)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0\,,$$

defined on $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is the “copy space”.

Optimal IS: the decoupling algorithm

We compute the optimal importance sampling measure for SDE

$$\mathrm{d}\overline{X}_t = b(t, \overline{X}_t, \mu_t^N)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0\,,$$

defined on $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is the “copy space”.

Goal: minimize over $h \in \mathbb{H}_T$ the variance conditional on μ^N:

$$\mathbb{E}_{\mathbb{P}\otimes\tilde{\mathbb{P}}}\big[\, G(\overline{X}_T)^2 Z_T^{-1} \big| \tilde{\mathcal{F}}_T\big], \quad \mathrm{d}Z_t = \dot{h}_t Z_t \mathrm{d}W_t^{\mathbb{P}}, \quad Z_0 = 1.$$

Optimal IS: the decoupling algorithm

We compute the optimal importance sampling measure for SDE

$$\mathrm{d}\overline{X}_t = b(t, \overline{X}_t, \mu^N_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0\,,$$

defined on $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is the “copy space”.

Goal: minimize over $h \in \mathbb{H}_T$ the variance conditional on μ^N:

$$\mathbb{E}_{\mathbb{P}\otimes\tilde{\mathbb{P}}}\big[\, G(\overline{X}_T)^2 Z_T^{-1} \big| \tilde{\mathcal{F}}_T \big], \quad \mathrm{d}Z_t = \dot{h}_t Z_t \mathrm{d}W^{\mathbb{P}}_t, \quad Z_0 = 1.$$

Small-noise asymptotics:

$$L(h;\mu^N) = \limsup_{\epsilon\to 0} \epsilon \log \mathbb{E}_{\mathbb{P}\otimes\tilde{\mathbb{P}}}\left[\exp\left(\frac{1}{\epsilon}\Big(2\log(\overline{G}(\sqrt{\epsilon}W)) - \int_0^T \sqrt{\epsilon}\dot{h}_t \mathrm{d}W_t + \frac{1}{2}\int_0^T \dot{h}_t^2 \mathrm{d}t\Big)\right) \Big| \tilde{\mathcal{F}}_T\right]$$

Optimal IS: the decoupling algorithm

We compute the optimal importance sampling measure for SDE

$$\mathrm{d}\overline{X}_t = b(t, \overline{X}_t, \mu_t^N)\mathrm{d}t + \sigma \mathrm{d}W_t, \qquad X_0 = x_0\,,$$

defined on $(\Omega, \mathcal{F}, \mathbb{P}) \otimes (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ where $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ is the "copy space".

Goal: minimize over $h \in \mathbb{H}_T$ the variance conditional on μ^N:

$$\mathbb{E}_{\mathbb{P}\otimes\tilde{\mathbb{P}}}\big[\, G(\overline{X}_T)^2 Z_T^{-1} \big| \tilde{\mathcal{F}}_T\big], \quad \mathrm{d}Z_t = \dot{h}_t Z_t \mathrm{d}W_t^{\mathbb{P}}, \quad Z_0 = 1.$$

Small-noise asymptotics:

$$L(h;\mu^N) = \limsup_{\epsilon\to 0} \epsilon \log \mathbb{E}_{\mathbb{P}\otimes\tilde{\mathbb{P}}}\left[\exp\left(\frac{1}{\epsilon}\Big(2\log(\overline{G}(\sqrt{\epsilon}W)) - \int_0^T \sqrt{\epsilon}\dot{h}_t \mathrm{d}W_t + \frac{1}{2}\int_0^T \dot{h}_t^2 \mathrm{d}t\Big)\right)\Big|\tilde{\mathcal{F}}_T\right]$$

A change of measure parameter $h^* \in \mathbb{H}_T$ is asymptotically optimal w.r.t. μ^N if

$$L(h^*;\mu^N) = \min_{h\in\mathbb{H}_T} L(h;\mu^N) \quad \tilde{\mathbb{P}}\text{-a.s.}\,.$$

Optimal IS: the decoupling algorithm

Under above assumptions, the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$L(h;\mu^N) = \sup_{u \in \mathbb{H}_T} \left\{ 2\log(\overline{G}(u)) - \int_0^T \dot{h}_t \dot{u}_t \mathrm{d}t + \frac{1}{2}\int_0^T \dot{h}_t^2 \mathrm{d}t - \frac{1}{2}\int_0^T \dot{u}_t^2 \mathrm{d}t \right\} \quad \tilde{\mathbb{P}}\text{-a.s.}.$$

Optimal IS: the decoupling algorithm

Under above assumptions, the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$L(h; \mu^N) = \sup_{u \in \mathbb{H}_T} \left\{ 2 \log(\overline{G}(u)) - \int_0^T \dot{h}_t \dot{u}_t \mathrm{d}t + \frac{1}{2} \int_0^T \dot{h}_t^2 \mathrm{d}t - \frac{1}{2} \int_0^T \dot{u}_t^2 \mathrm{d}t \right\} \quad \tilde{\mathbb{P}}\text{-a.s.} .$$

ii. There exists an asymptotically optimal h^* which minimizes $L(h; \mu^N)$.

Optimal IS: the decoupling algorithm

Under above assumptions, the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$L(h;\mu^N) = \sup_{u\in\mathbb{H}_T}\left\{2\log(\overline{G}(u)) - \int_0^T \dot{h}_t\dot{u}_t\mathrm{d}t + \frac{1}{2}\int_0^T \dot{h}_t^2\mathrm{d}t - \frac{1}{2}\int_0^T \dot{u}_t^2\mathrm{d}t\right\} \quad \tilde{\mathbb{P}}\text{-a.s.}\,.$$

ii. There exists an asymptotically optimal h^* which minimizes $L(h;\mu^N)$.

iii. Consider a simplified optimization problem

$$\sup_{u\in\mathbb{H}_T}\left\{2\log(\overline{G}(u)) - \int_0^T \dot{u}_t^2\mathrm{d}t\right\}.$$

There exists a maximizer h^{**} for this problem. If

$$L(h^{**};\mu^N) = 2\log(\overline{G}(h^{**})) - \int_0^T (\dot{h}_t^{**})^2\mathrm{d}t\,,$$

then h^{**} is asymptotically optimal.

Optimal IS: complete measure change

We want to minimize $\mathbb{E}_{\mathbb{P}}\left[G(X)^2 \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right]$. Consider a particle approximation of X:

$$\mathrm{d}X_t^{i,N} = b\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}}\right)\mathrm{d}t + \sigma \mathrm{d}W_t^{i,\mathbb{P}}, \quad X_0^{i,N} = x_0,$$

$$\mathrm{d}Z_t^i = \dot{h}_t Z_t^i \mathrm{d}W_t^{i,\mathbb{P}}, \quad \mathcal{E}_0^i = 1,$$

and minimize $\mathbb{E}_{\mathbb{P}}[G^2(X^{i,N})(Z_T^i)^{-1}]$ over all $h \in \mathbb{H}_T$.

Optimal IS: complete measure change

We want to minimize $\mathbb{E}_{\mathbb{P}}\left[G(X)^2 \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right]$. Consider a particle approximation of X:

$$\mathrm{d}X_t^{i,N} = b\left(t, X_t^{i,N}, \frac{1}{N}\sum_{j=1}^{N}\delta_{X_t^{j,N}}\right)\mathrm{d}t + \sigma \mathrm{d}W_t^{i,\mathbb{P}}, \quad X_0^{i,N} = x_0\,,$$

$$\mathrm{d}Z_t^i = \dot{h}_t Z_t^i \mathrm{d}W_t^{i,\mathbb{P}}, \quad \mathcal{E}_0^i = 1,$$

and minimize $\mathbb{E}_{\mathbb{P}}[G^2(X^{i,N})(Z_T^i)^{-1}]$ over all $h \in \mathbb{H}_T$.

Small-noise asymptotics:

$$\bar{L}(h) := \limsup_{\epsilon\to 0} \epsilon \log\left(\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{1}{\epsilon}\Big(2\log\big(\tilde{G}_i(\sqrt{\epsilon}W^1, \ldots, \sqrt{\epsilon}W^N)\big) - \int_0^T \sqrt{\epsilon}\dot{h}_t \mathrm{d}W_t^i + \frac{1}{2}\int_0^T (\dot{h}_t)^2 \mathrm{d}t\Big)\right)\right]\right), \quad h \in \mathbb{H}_T$$

where, $\tilde{G}_i(W^1, \ldots, W^N) := G(X^{i,N}(W^1, \ldots, W^N))$.

Optimal IS: complete measure change

Under the above assumptions the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$\bar{L}(h) = \sup_{u \in \mathbb{H}_T^N} \left\{ 2\log(\tilde{G}_1(u^1, \dots, u^N)) - \int_0^T \dot{h}_t \dot{u}_t^1 \mathrm{d}t + \frac{1}{2}\int_0^T (\dot{h}_t)^2 - |\dot{u}_t|^2 \mathrm{d}t \right\}.$$

Optimal IS: complete measure change

Under the above assumptions the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$\bar{L}(h) = \sup_{u \in \mathbb{H}_T^N} \left\{ 2 \log(\tilde{G}_1(u^1, \ldots, u^N)) - \int_0^T \dot{h}_t \dot{u}_t^1 \mathrm{d}t + \frac{1}{2} \int_0^T (\dot{h}_t)^2 - |\dot{u}_t|^2 \mathrm{d}t \right\}.$$

ii. There exists an asymptotically optimal parameter h^* minimizing $\bar{L}(h)$.

Optimal IS: complete measure change

Under the above assumptions the following statements hold:

i. Let $h \in \mathbb{H}_T$ such that $\dot{h}$ is of finite variation. Then

$$\bar{L}(h) = \sup_{u \in \mathbb{H}_T^N} \left\{ 2\log(\tilde{G}_1(u^1, \dots, u^N)) - \int_0^T \dot{h}_t \dot{u}_t^1 \mathrm{d}t + \frac{1}{2}\int_0^T (\dot{h}_t)^2 - |\dot{u}_t|^2 \mathrm{d}t \right\}.$$

ii. There exists an asymptotically optimal parameter h^* minimizing $\bar{L}(h)$.

iii. Consider a simplified optimization problem

$$\sup_{u^1 \in \mathbb{H}_T, \hat{u} \in \mathbb{H}_T} \left\{ 2\log(\tilde{G}_1(u^1, \hat{u}, \dots, \hat{u})) - \int_0^T (\dot{u}_t^1)^2 \mathrm{d}t - \frac{N-1}{2}\int_0^T \dot{\hat{u}}_t^2 \mathrm{d}t \right\}.$$

There exists a maximizer (h^{**}, u^{**}) for this problem. If

$$\bar{L}(h^{**}) = 2\log\left(\tilde{G}_1(h^{**}, u^{**}, \dots, u^{**})\right) - \int_0^T (\dot{h}_t^{**})^2 \mathrm{d}t - \frac{N-1}{2}\int_0^T (\dot{u}_t^{**})^2 \mathrm{d}t .$$

then h^{**} is asymptotically optimal.

Outline

Example: the Kuramoto model

Consider the Kuramoto model used to study various phenomena in physics:

$$\mathrm{d}X_t = \left(K \int_{\mathbb{R}} \sin(y - X_t)\mu^X_{t,\mathbb{P}}(\mathrm{d}y) - \sin(X_t) \right) \mathrm{d}t + \sigma \mathrm{d}W^{\mathbb{P}}_t \,, \quad t \in [0, T], \;\; X_0 = x_0 \,,$$

where K is the coupling strength and σ has the physical interpretation of the temperature in the system.

Our **goal** is to obtain the asymptotically optimal change of measure that improves the estimation of $\mathbb{E}_{\mathbb{P}}[G(\bar{X}_T)]$.

Obs: Intentionally we do not split $\sin(y - x) = \cos(x)\sin(y) - \sin(x)\cos(y)$ as in *Bencheikh Jourdain 2018*

Example 1: regular terminal condition

Terminal condition $G(x) = a\exp(bx)$ and use $T = 1$, $\bar{X}_0 = 0$, $K = 1$, $\sigma = 0.3$, $a = 0.5$ and $b = 10$ with an Euler scheme with step size $\Delta t = 0.02$.

	Monte Carlo			Decoupled			Complete		
N	Payoff	Std.Err.	Time	Payoff	Std.Err.	Time	Payoff	Std.Err.	Time
1×10^3	1.5066	0.1490	3	1.5729	0.0028	9	1.5419	0.0024	3
5×10^3	1.5895	0.0626	27	1.5840	0.0013	54	1.5710	0.0013	28
1×10^4	1.6813	0.0693	76	1.5728	0.0009	153	1.5860	0.0009	75
5×10^4	1.5899	0.0200	1 025	1.5820	0.0004	2 052	1.5738	0.0004	1 062
1×10^5	1.5807	0.0176	3 433	1.5731	0.0003	6 935	1.5882	0.0003	3 644

Observation:

- **Time**: One can truly observe the $O(N^2)$ scaling (Hence hard to reduce the error using standard MC). The decoupling takes double the time.
- **Error**: The variance coming from the IS schemes are vastly smaller then that of standard Monte Carlo: 2-3 orders of magnitude
- **Prop. of Chaos error:** For standard Monte Carlo, with $N = 5 \times 10^3$, the propagation of chaos error is ≈ 0.0041.
- **Payoff**: The two IS schemes are slightly off each other. Possibly the measure change is changing the particle distribution too much.

Example 2: a terminal condition with steep slope

To mimic a rare event, consider the terminal condition

$$G(x) = \big(\tanh(a(x-b)) + 1\big)/2$$

for a large (mollified indicator function). We take $a = 15$ and $b = 1$.

All values in the table are to be multiplied by 10^{-9}.

	Monte Carlo		Decoupled		Complete	
N	Payoff	Std.Err.	Payoff	Std.Err.	Payoff	Std.Err.
1×10^3	1.015	0.671	3.864	0.0250	8.456	0.101
5×10^3	1.093	0.752	3.952	0.0112	5.564	0.0185
1×10^4	8.829	7.071	3.910	0.0077	32.956	0.1520
5×10^4	1.106	0.271	3.970	0.0035	2.101	0.0024
1×10^5	5.158	1.990	3.901	0.0024	16.781	0.019

The complete measure change creates instability by introducing error into the estimation of coefficients

Conclusions

- We propose two importance sampling algorithms for MV-SDE: decoupling and complete measure change.
  - The decoupling alg. requires twice as many simulations to estimate the law of the solution.
  - For not so tough problems, both methods have similar performance, and strongly reduce the variance.
- For very rare events, complete measure change may introduce instability and the decoupling algorithm is preferred.
- **Key remaining question**:
  - Extend to the case of $\mathbb{E}[G(X_T, \mu_T)]$

Thank you!

Thank you for your time!

Some References

- GdR, G. Smith and P. Tankov, *Importance Sampling for McKean-Vlasov SDEs*, 2018, arXiv:1803.09320
- GdR, S. Engelhardt and G. Smith, *Simulation of McKean Vlasov SDEs with super linear growth*, 2018, arXiv:1808.05530,
- GdR, W. Salkeld and J. Tugaut, *Freidlin-Wentzell LDPs in path space for McKean-Vlasov equations and the Functional Iterated Logarithm Law*, 2017, arXiv:1708.04961 - To appear in Annals of Applied Probability

- Glasserman, Heidelberger and Shahabuddin *Asymptotically Optimal Importance Sampling and Stratification for Pricing Path Dependent Options*, 1999
- P. Guasoni and S. Robertson, *Optimal importance sampling with explicit formulas in continuous time*, Finance and Stochastics 12 (2008), no. 1, 1–19
- A. Kohatsu-Higa and S. Ogawa, *Weak rate of convergence for an Euler scheme of nonlinear SDE's*, Monte Carlo Methods and Applications 3 (1997), 327–345.