Topics in Retrospective Simulation

Gareth Roberts

Department of Statistics University of Warwick

Part 2: Rao-Blackwellisation for Improved Monte Carlo for Stochastic Processes

Including work with Alex Beskos, Paul Fearnhead, Krys Latuszynski, Omiros Papaspiliopoulos, Giorgos Sermaidis and Sylvain Le Corff

Plan for presentation

1. Introduction to Rao-Blackwellisation for Monte Carlo
2. Retrospective Rejection Sampling
3. Exact simulation of diffusions
4. Rao-Blackwellised unbiased importance sampling for stochastic processes

Plan for presentation

1. Introduction to Rao-Blackwellisation for Monte Carlo
2. Retrospective Rejection Sampling
3. Exact simulation of diffusions
4. Rao-Blackwellised unbiased importance sampling for stochastic processes

The Rao-Blackwell Theorem

If $\hat{\theta}$ is some estimator of θ, then $\mathbf{E}(\hat{\theta}|S)$ for any sufficient statistic S is at least as good (in the sense of mean square error).

S being sufficient is required only to ensure $\mathbf{E}(\hat{\theta}|S)$ is a function of the data alone and not the unknown parameter θ.

Rao-Blackwell for Markov chain Monte Carlo (MCMC)

Suppose $\{X_n\}$ is a positive recurrent Markov chain with invariant probability measure π, and we wish to estimate $\pi(f) = \mathbf{E}_\pi(f(X))$. A natural estimator is

$$E_1 = \frac{\sum_{n=1}^{N} f(X_n)}{N} .$$

If $f \in L^1(\pi)$ this is a consistent estimator and commonly can be shown to satisfy a central limit theorem.

However suppose we can analytically calculate $g(X_n) = \mathbf{E}(f(X_{n+1})|X_n)$, then by the Rao-Blackwell Theorem,

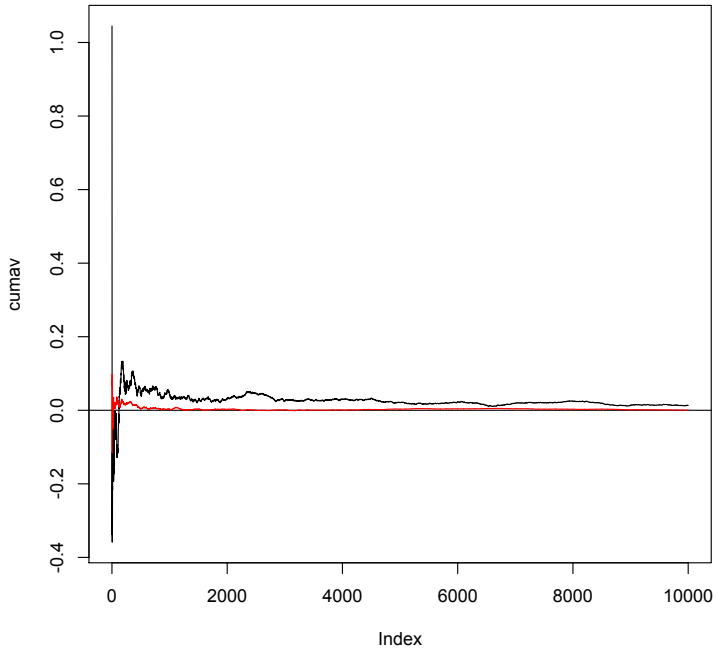
$$\mathrm{Var}(g(X_n)) \leq \mathrm{Var}(f(X_n))$$

suggesting instead we use the estimator:

$$E_2 = \frac{\sum_{n=1}^{N} g(X_n)}{N}$$

Rao-Blackwellisation for MCMC output

See in particular Robert and Casella (1996).



The ergodic average using E_1 and E_2. In general, the variance gain can be arbitrarily large, but is usually quite small.

In fact Rao-Blackwellised estimator CAN be worse (Liu Wong King 1994), though there are positive results for reversible MCs (McKeague and Wefelmeyer, 2000)

Later we shall see an example where Rao-Blackwellisation is necessary even for the estimator to exist.

Rao-Blackwellised Kernel density estimates

Consider a d-dimensional Markov chain $(\theta_n^{(1)}, \ldots \theta_n^{(d)})$, $n = 0, 1, 2, \ldots$, and let $\pi^{(1)}(\theta^{(1)}|\theta^{(-1)})$ denote the full conditional of the first component given all the others.

Rao-Blackwellised kernel density estimate for $\theta^{(1)}$ is

$$K(\theta^{(1)}) = \frac{\sum_{n=1}^{N} \pi^{(1)}(\theta^{(1)}|\theta_n^{(-1)})}{N}$$

Robert and Casella (1996)

Note that this is a Rao-Blackwellised estimator assuming we are running a Gibbs sampler which uses $\theta^{(1)}$ as one of its constituent steps.

However we can still use this estimator for ANY Markov chain with invariant distribution π.

Plan for presentation

1. Introduction to Rao-Blackwellisation for Monte Carlo
2. Retrospective Rejection Sampling
3. Exact simulation of diffusions
4. Rao-Blackwellised unbiased importance sampling for stochastic processes

Rejection sampling

Let f be a density of interest, and g be a density from which we can simulate. f/g bounded by K say.

1. Sample X from g.
2. Compute $p(X) = f(X)/(Kg(X))$.
3. Simulate $U \sim U(0, 1)$.
4. Accept X if $p(X) > U$. Otherwise return to 1.

Blue steps are often unnecessary!

Retrospective rejection sampling

1. Sample $V \sim U(0,1)$.
2. Identify a function $h(V,X)$ and a set $A(V)$ such that

$$\mathbf{P}_V\{h(V,X) \in A(V)\} = p(X)$$

3. Simulate $h(X,V)$.
4. If $h(X,V) \in A(V)$ the accept. Otherwise return to 1.
5. Fill in missing bits of X from distribution of $X|h(X,V)$ as required.

Simulation of stochastic processes

Suppose that $X : [0,1] \to \mathbf{R}^d$ is a stochastic process with associated probabiltiy measure $\mathbf{P}_0$.

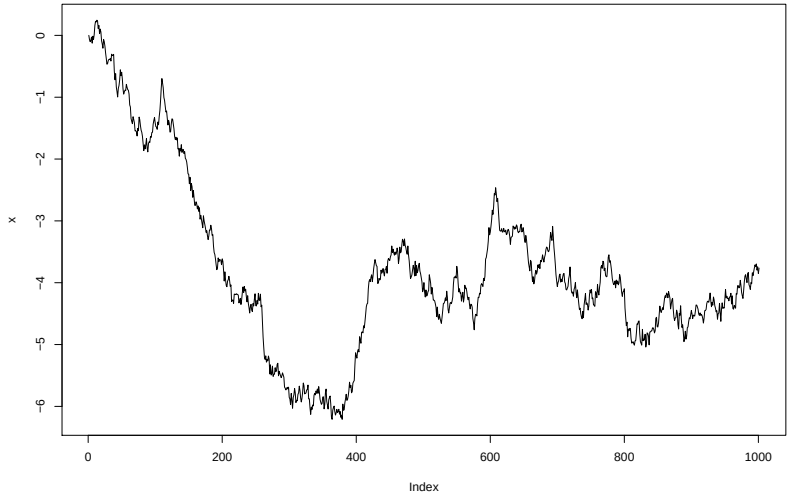
Suppose we are able to simulate from $\mathbf{P}_0$.

Suppose that we wish to simulate from a different distribution $\mathbf{P}$ which cannot be directly simulated, but for which we can write:

$$\frac{d\mathbf{P}}{d\mathbf{P}_0}(X) \propto \exp\{-r \int_0^1 \phi(X_s) ds\} = a(X)$$

for some function ϕ taking values in $[0,1]$.

This applies to very wide range of stochastic processes, eg point processes in space and time, diffusions, jump diffusions, processes used in Bayesian non-parametrics.



For example, given this trajectory, $a(X)$ describes the Radon-Nikodym derivative between $\mathbf{P}$ and $\mathbf{P}_0$ for this particular trajectory.

Rejection for sample paths

Would like to just propose a sample path fom $\mathbf{P}_0$ and use rejection sampling. However

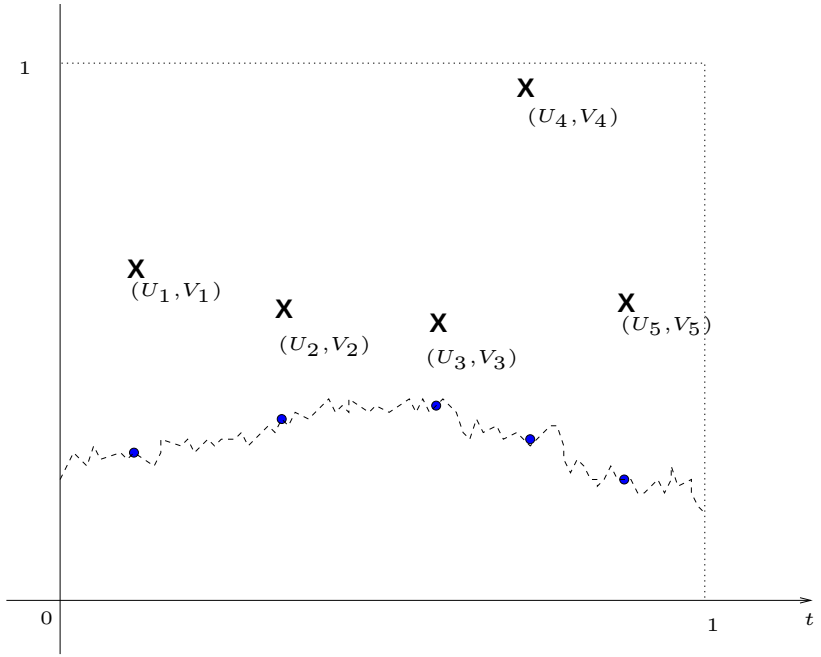
- Just storing all of X could require infinite storage capacity.
- Calculating $\int_0^1 \phi(X_s)ds$ is likely to require infinite computation

We could approximate in some way, but this seems unsatisfactory, and it would typically be very difficult to quantify the resulting approximation error.

Retrospective rejection simulation

Key observation: $a(x)$ is the probability of a Poisson random variable of parameter $r \int_0^1 \phi(X_s)ds$ taking value 0.

Or ... the probability that a Poisson process of rate r on the unit square has no points on the epi graph $\{(u, v) \in [0, 1]^2;\ v \leq \phi(u)\}$.



Plan for presentation

1. Introduction to Rao-Blackwellisation for Monte Carlo
2. Retrospective Rejection Sampling
3. Exact simulation of diffusions
4. Rao-Blackwellised unbiased importance sampling for stochastic processes

Simulation of diffusions

Continuous, strong Markov processes described by stochastic differential equation:

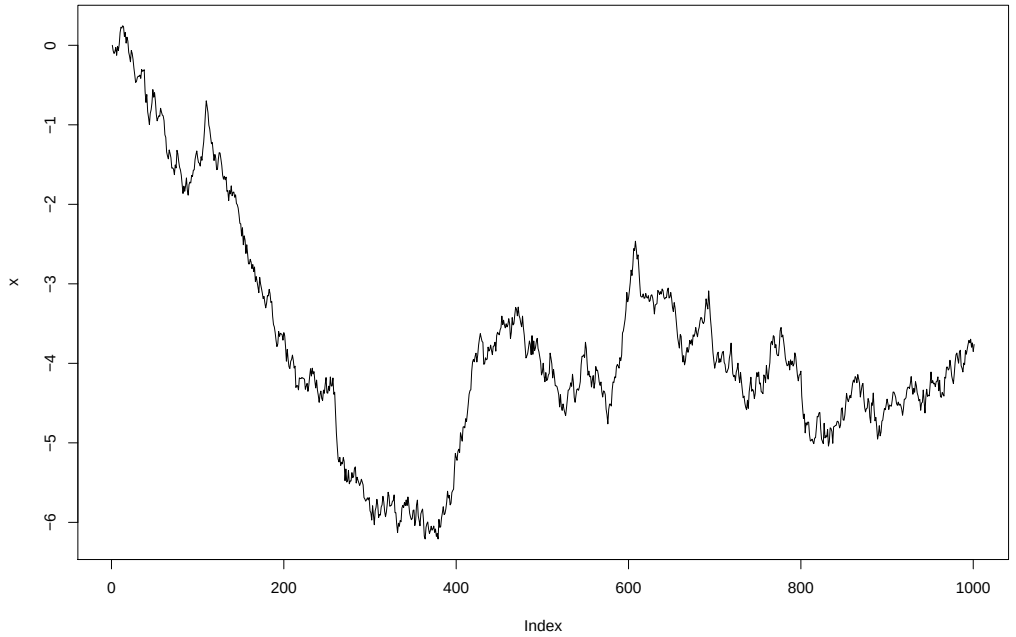
$$dX_t = \alpha(X_t)dt + \sigma(X_t)dB_t$$

where B is standard Brownian motion.

This can be interpreted constructively as

$$X_{t+\epsilon} = X_t + \epsilon\alpha(X_t) + \sigma(X_t)N(0, \epsilon)$$

approximately for 'small' ϵ (the **Euler approximation**) written as



Interested in simulating without discretisation error and obtaining a realisation of the whole path in some sense.

Diffusion densities

Consider simplest case, σ constant and drift α which is bounded with bounded derivative.

$$dX_t = \alpha(X_t)dt + dB_t$$

and let the law of this diffusion on $[0,1]$ be denoted $\mathbf{P}$, with $\mathbf{W}$ being that of the Brownian motion (Wiener measure) .

Then under very weak regularity conditions

$$\frac{d\mathbf{P}}{d\mathbf{W}}(X) = G(X)$$

where G is given by the **Cameron-Martin-Girsanov** formula:

$$\log G(X) = \int_0^1 \left(\alpha(X_s)dX_s - \alpha^2(X_s)/2\right) ds$$

Towards a simulation algorithm: simplifying G

By a suitable rearrangement we can rewrite

$$\frac{d\mathbf{P}}{d\mathbf{W}}(X) = G(X) \propto \exp\left\{ A(X_1) - r\int_0^1 \phi(X_s)ds \right\} := a(X)$$

where ϕ always always takes values in the interval $[0, 1]$.

This is almost in the exponential form required for the Poisson process idea above.

So we consider biased Brownian motion proposals for rejection sampling:

$$\mathbf{P}_0(X_1 \in dx) \propto \exp\{A(x) - x^2/2\}\, dx \qquad (*)$$

with $\mathbf{X}|X_1 \sim$Brownian bridge, so that

$$\frac{d\mathbf{P}}{d\mathbf{P}_0} \propto \exp\left\{-r \int_{s=0}^{1} \phi(X_s) ds\right\}.$$

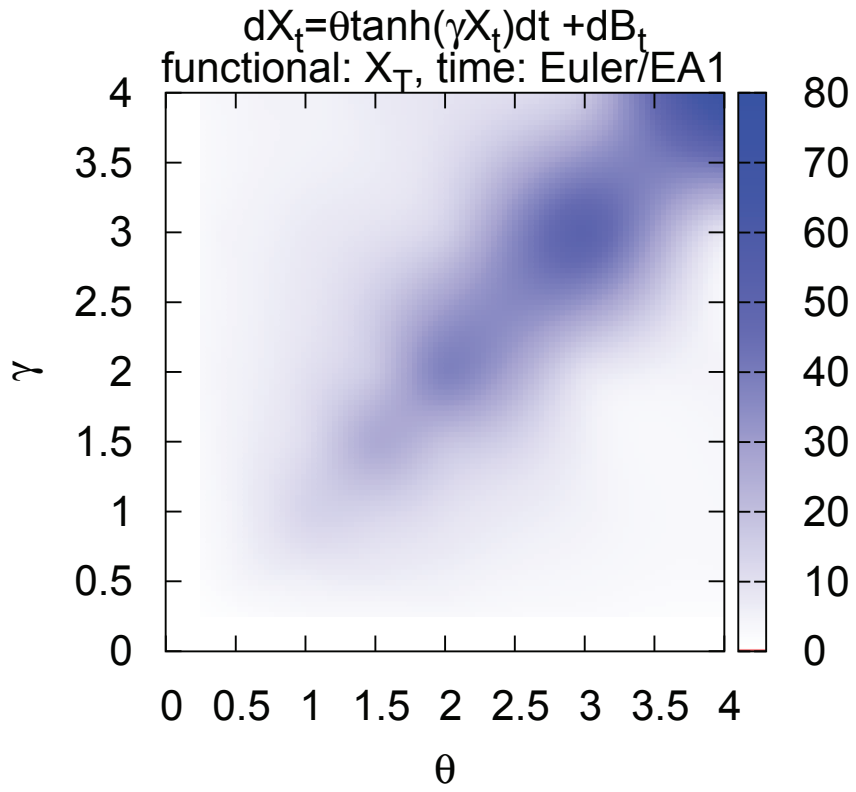
Let Φ be a Poisson process of rate r on $\{0 \le y \le \phi(X_s),\ 0 \le s \le 1\}$. Then

$$\mathbf{P}\left(\Phi \text{ is the empty configeration} = \exp\left\{-r \int_{0}^{1} \phi(X_s) ds\right\}\right).$$

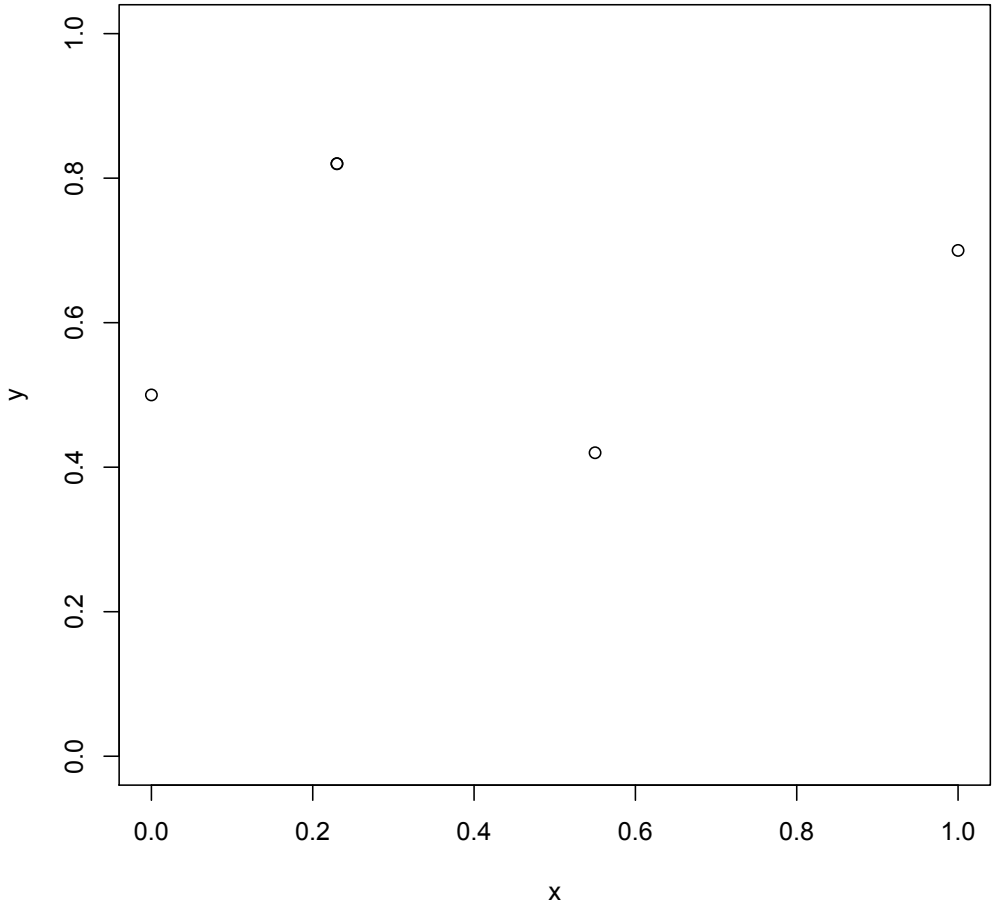
The basic diffusion Exact Algorithm (EA1)

1. Set $B_0 = 0$. Simulate B_1 from (*)
2. Generate Poisson process of rate r on $[0,1] \times [0,1]$: $\Phi = \{(U_1, V_1), \ldots (U_n, V_n)\}$
3. For each U_i, draw B_{U_i} from its appropriate Brownian bridge probabilities.
4. If $\phi(B_{U_i}) > V_i$ for ANY i, erase skeleton and go to (1).
5. Output the currently stored skeleton $\{(0, B_0), (1, B_1), (U_i, B_{U_i}),\ 1 \leq i \leq n\}$.

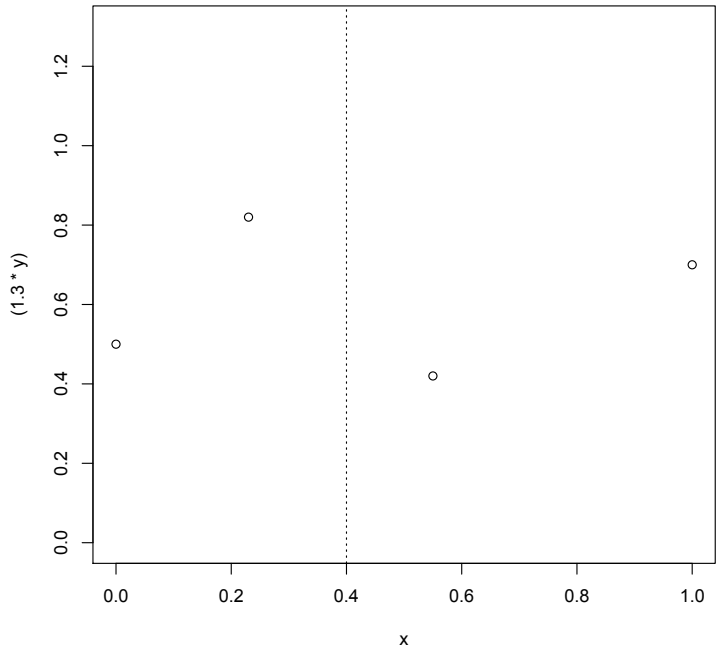
Part of a simulation study



The algorithm output



How can we use the output for Monte Carlo?



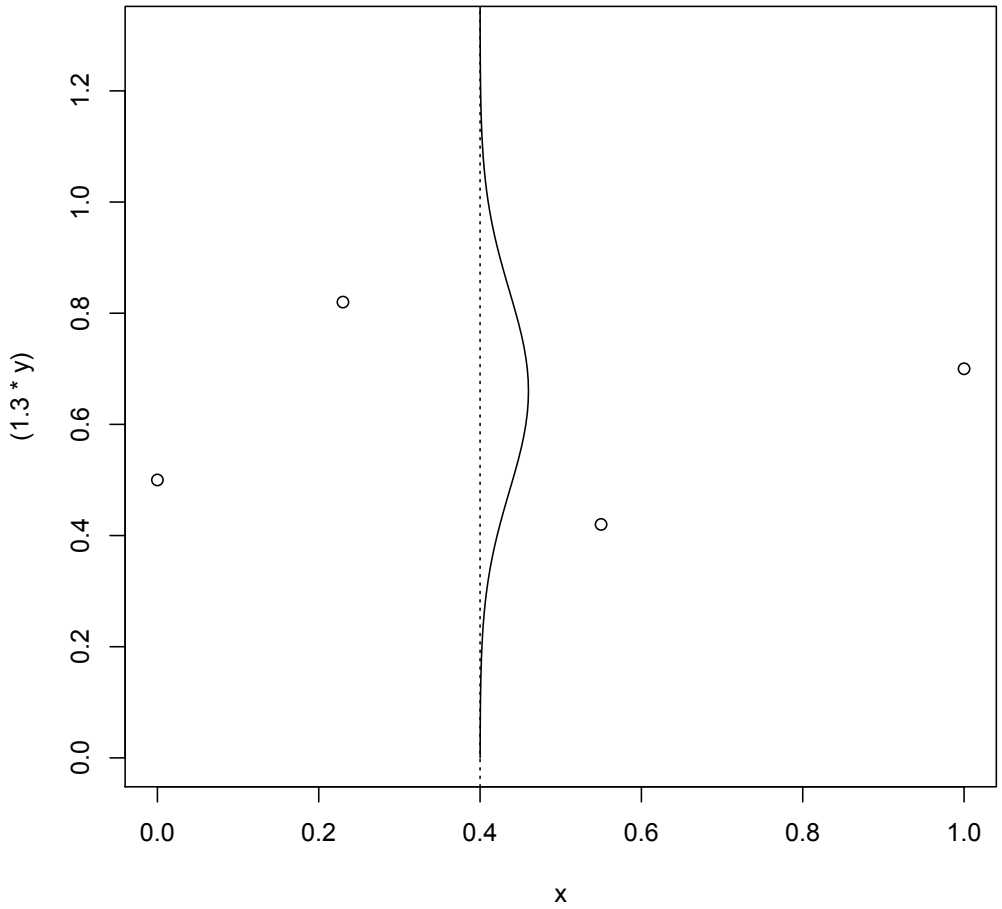
This is the skeleton $S(X)$, a finite and random dimensional collection of times together with the value of X at those times:

$$\{U_0 = 0, U_1, \ldots, U_{\kappa+1} = 1; X_0, X_{U_1}, \ldots, X_{U_\kappa}, X_1\}$$

We can discard the Poisson process heights, $V_1, \ldots V_\kappa$.

Distribution of X_t not contained in S

Suppose we wish to estimate $\mathbf{E}(X_{.4}^3)$.



Random Sufficiency

$$dX_t = \alpha(X_t)dt + dB_t$$

Let $\tilde{S}$ be a proposed skeleton

$$\{U_0 = 0, U_1, \ldots, U_{\kappa+1} = 1; X_0, X_{U_1}, \ldots, X_{U_\kappa}, X_1\}$$

Set D denote the accept reject decision:

$$D = \prod_{i=1}^{\kappa} \mathbf{1}_{\phi(X_{U_i}) \leq V_i} \ .$$

Then X and D are independent conditional on $\tilde{S}$, so that the distribution of $X|\tilde{S}$ is independent of whether the proposal is accepted.

Since the distribution of $X|X_1$ under the proposal measure is independent of α, it follows that S acts as a random sufficient statistic for α in the sense that the distribution of $X|S$ does not depend on α.

In fact $X|S$ can be constructed as a collection of $\kappa + 1$ Brownian bridges linking successive points in S.

Estimation

Suppose we have N skeletons, $S_1, \ldots S_N$. Conditional on S, $X_{0.4} \sim N(a(S), b(S))$ for some a, b according to the appropriate Brownian bridge probabilities for the bridge bridging time 0.4. So independent exact draws from $X_{0.4}$ can be obtained by simulating

$$X_i \sim N(a(S_i), b(S_i)), \quad 1 \leq i \leq N \ .$$

$$E_1 = \frac{\sum_{n=1}^{N} X_i^3}{N} \ .$$

But $\mathbf{E}(X^3|S)$ $(= c(S)$ say) is analytically tractable, so that the Rao-Blackwellised estimator

$$E_2 = \frac{\sum_{n=1}^{N} c(S_i)}{N} \ .$$

will have smaller variance.

How to do with state-dependent volatility?

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt$$

We can apply the Lamperti transform: $Y_t = h(X_t)$ where

$$h(x) = \int_0^x \frac{dz}{\sigma(z)}$$

By Itô's formula we get

$$dY_t = dB_t + \alpha(Y_t)$$

where

$$\alpha(y) = \frac{b(h^{-1}(y))}{\sigma(h^{-1}(y))} - \frac{\sigma'(h^{-1}(y))}{2} .$$

Plan for presentation

1. Introduction to Rao-Blackwellisation for Monte Carlo
2. Retrospective Rejection Sampling
3. Exact simulation of diffusions
4. Rao-Blackwellised unbiased importance sampling for stochastic processes

The Exact Algorithm for multi-dimensional diffusions

Generally simulation and inference for diffusions is performed by approximating the diffusions by a discrete-time Markov process.

What about multiple-dimensional problems? Now σ is matrix, b a vector etc.

$$dX_t = \sigma(X_t)dB_t + b(X_t)dt$$

Firstly, can we reduce to the constant volatility case as we did on 1-dimension?

In general we cannot do this ...

The multi-dimensional diffusion case

For multi-dimensional diffusions, we can adopt the exact algorithm if:

- The volatility can be transformed to be constant via the Lamperti transform: ie we can find a 1-1 function η satisfying the matrix valued differential equation

$$(\nabla \eta)\sigma = I_d$$

- The drift of the transformed diffusion is the gradient of a potential: $\mu(x) = \nabla A(x)$.

This can be applied to almost all 1-d diffusions for which CMG theorem holds, but only certain classes of d-dimensional ones.

Why?

The exact Algorithm is a Rejection Sampler based on proposing paths from a drift-less version of the diffusion (with same volatility).

The acceptance probability for the path is (for $\sigma(x) = I_d$) proportional to:

$$\exp\left\{\int_0^T \mu(X_t)\mathrm{d}X_t - \frac{1}{2}\int_0^T |\mu(X_t)|^2\mathrm{d}t\right\}$$
$$= \exp\left\{A(X_T) - A(X_0) - \frac{1}{2}\int_0^T \left(|\mu(X_t)|^2 + \nabla\mu(X_t)\right)\mathrm{d}t\right\}.$$

Whilst this cannot be evaluated, events with this probability can be simulated.

The condition $\mu(x) = \nabla A(x)$ is required to replace the stochastic integral by a Lebesgue one. It is a necessary and sufficient condition for Girsanov's formula to be bounded for bounded sample paths.

The condition $\mu(x) = \nabla A(x)$ is required to replace the stochastic integral by a Lebesgue one. It is a necessary and sufficient condition for Girsanov's formula to be bounded for bounded sample paths.

The condition $\sigma(x)$ is constant is so that we can simulate from the driftless diffusion.

The condition $\mu(x) = \nabla A(x)$ is required to replace the stochastic integral by a Lebesgue one. It is a necessary and sufficient condition for Girsanov's formula to be bounded for bounded sample paths.

The condition $\sigma(x)$ is constant is so that we can simulate from the driftless diffusion.

- Importance sampling seems doomed if we cannot sample from an distribution wrt which target is absolutely continuous.

Consider two diffusions with different diffusion coefficients, σ_1 and σ_2, then their laws as NOT mutually absolutely continuous ...

even though their finite-dimensional distributions typically are.

Transition Densities

We will denote the transition density of the diffusion by

$$p(y|x,h)dy = p(X_{t+h} \in dy | X_t = x)$$

.

It satisfies Kolmogorov's forward equation:

$$\frac{\partial}{\partial t} p(y|x,t) = \mathcal{K}_y p(y|x,t),$$

for some forward-operator $\mathcal{K}_y$ which acts on y.

Generally the transition density is intractable with the usual exceptions: constant or linear drifts and a few others ...

Avoiding time-discretisation Errors: Why?

Why might time-discretisation error free methods be useful?

- Error in estimates are purely Monte Carlo. Thus it is easier to quantify the error.

Avoiding time-discretisation Errors: Why?

Why might time-discretisation error free methods be useful?

- Error in estimates are purely Monte Carlo. Thus it is easier to quantify the error.
- Time-discretisation may tend to use substantially finer discretisations than are necessary: possible computational gains?

Avoiding time-discretisation Errors: Why?

Why might time-discretisation error free methods be useful?

- Error in estimates are purely Monte Carlo. Thus it is easier to quantify the error.
- Time-discretisation may tend to use substantially finer discretisations than are necessary: possible computational gains?
- Want methods which are robust as $h \to 0$

Avoiding time-discretisation Errors: Why?

Why might time-discretisation error free methods be useful?

- Error in estimates are purely Monte Carlo. Thus it is easier to quantify the error.
- Time-discretisation may tend to use substantially finer discretisations than are necessary: possible computational gains?
- Want methods which are robust as $h \to 0$
- Error is $O(C^{-1/2})$, where C is CPU cost. Alternative approaches have errors that can be e.g. $O(C^{-1/3})$ (though see multigrid work by Giles and others).

Generalising the exact algorithm using importance sampling

Our aim was to try and extend the ability to perform simulation / estimation / inference without time-discretisation approximations to a wider class of diffusions.

The key is to be able to unbiasedly estimate expectations, such as $\mathrm{E}(f(X_t))$ or $\mathrm{E}(f(X_{t_1}, \ldots, X_{t_m}))$.

The approach we have developed can be applied to general continuous-time Markov processes, and is a continuous-time version of sequential importance sampling.

We construct a signed measure-valued stochastic processes (which is non-Markov) $\{\xi_t, t \geq 0\}$ with

$$\mathrm{E}(\xi_t(f)) = \mathrm{E}(f(X_t))$$

Unbiased estimation almost as good as simulation given recent advances in combining importance sampling and simulation algorithms such as MCMC.

Importance Sampling

Importance Sampling (**IS**) is a Monte Carlo integration technique. Consider the integral

$$I = \int f(x)p(x)\mathrm{d}x = \int w(x)f(x)q(x)\mathrm{d}x,$$

where $p(x)$ and $q(x)$ are densities, $f(x)$ is arbitrary and $p(x) > 0 \Rightarrow q(x) > 0$. Here we are setting $w(x) = p(x)/q(x)$.

We can view this as an expectation with respect to $q(x)$. Thus

1. Sample x_i, $i = 1, \ldots, N$, iid from $q(x)$;
2. Estimate the integral by the unbiased, consistent estimator:

$$\hat{I} = \frac{1}{N}\sum_{i=1}^{N} w(x_i)f(x_i)$$

Sequential Importance Sampling (SIS)

As this gives an estimate of the expectation of $f(X)$ for arbitrary functions f, we can think of the sample from $q(x)$, and the corresponding weights as giving an approximation to the distribution defined by $p(x)$.

This idea can be extended to Markov processes:

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=2}^{n} p(x_i \mid x_{i-1}).$$

With a proposal process defined by $q(x_1)$ and $q(x_i \mid x_{i-1})$.

Sequential Importance Sampling (SIS)

To obtain one weighted sample:

1. Simulate $X_1^{(i)}$ from $q(x_1)$; assign a weight $\tilde{w}_1^{(i)} = p(x_1)/q(x_1)$.
2. For $t = 2, \ldots, n$; simulate $X_t^{(i)}|x_{t-1}^{(i)}$ from $q(x_t|x_{t-1}^{(i)})$, and set

$$\tilde{w}_t^{(i)} = \tilde{w}_{t-1}^{(i)} \frac{p(x_t^{(i)}|x_{t-1}^{(i)})}{q(x_t^{(i)}|x_{t-1}^{(i)})}.$$

New Approach: CIS

We now derive a continuous-time importance sampling (CIS) procedure for unbiased inference for general continuous-time Markov models.

We will describe the CIS algorithm for generating a single realisation. So at any time t we will have x_t and w_t, realisations of random variables X_t, W_t such that

$$\mathrm{E}_p(f(X_t)) = \mathrm{E}_q(f(X_t)W_t).$$

The former expectation is wrt to the target diffusion, the latter wrt to CIS procedure.

We will use a proposal process with tractable transition density $q(x|y,t)$ (and forward-operator $\mathcal{K}_x^{(1)}$).

A discrete-time SIS procedure

First consider a discrete-time SIS method aimed at inference at times $h, 2h, 3h, \ldots,$.

(0) Fix x_0; set $w_0 = 1$, and $i = 1$.

(1) Simulate $X_{ih} = x_{ih}$ from $q(x_{ih}|x_{(i-1)h})$.

(2) Set

$$w_i = w_{i-1}\frac{p(x_{ih}|x_{(i-1)h}, h)}{q(x_{ih}|x_{(i-1)h}, h)}$$

(3) Let $i = i + 1$ and goto (1).

Problems: cannot calculate weights, and often the efficiency degenerates as $h \to 0$ for fixed T.

As $h \to 0$, where q and p are discetisations of absolutely continuous diffusions, the limit is given by Girsanov's formula.

We want it to work in the case where q and p are mutually singular also!

Random weight SIS

It is valid to replace the weight in the SIS procedure by a random variable whose expectation is equal to the weight.

A simple way to do this here is to define

$$r(y, x, h) = 1 + \left(\frac{p(y|x, h)}{q(y|x, h)} - 1 \right) \frac{1}{\lambda h},$$

and introduce a Bernoulli random variable U_i, with success probability λh.

Then

$$\frac{p(y|x, h)}{q(y|x, h)} = \mathrm{E} \left\{ (1 - U_i) \cdot 1 + U_i r(y, x, h) \right\}.$$

Random weight SIS

Now we can have a random weight SIS algorithm:

(0) Fix x_0; set $w_0 = 1$, and $i = 1$.

(1) Simulate $X_{ih} = x_{ih}$ from $q(x_{ih}|x_{(i-1)h})$.

(2) Simulate U_i. If $U_i = 1$ then set $w_i = w_{i-1} r(x_{ih}, x_{(i-1)h}, h)$, otherwise $w_i = w_{i-1}$.

(3) Let $i = i + 1$ and goto (1).

This is a less efficient algorithm than the previous one, but it enables us to now use two tricks: retrospective sampling and Rao-Blackwellisation.

Retrospective Sampling

We only need to update the weights at time-points where $U_i = 1$. At these points we need to simulate $X_{ih}, X_{(i-1)h}$ to calculate the new weights.

If j is the most recent time when $U_j = 1$, then the distribution of X_{ih} is given by $q(x_{ih}|x_{jh}, (i-j)h)$ (assuming time-homogeneity for simplicity).

Given x_{jh} and x_{ih} the conditional distribution of $X_{(i-1)h}$ is

$$q(x_{(i-1)h}|x_{jh}, x_{ih}) = \frac{q(x_{(i-1)h}|x_{jh}, (i-j-1)h)q(x_{ih}|x_{(i-1)h}, h)}{q(x_{ih}|x_{jh}, (i-j)h)}.$$

New SIS algorithm

Using these ideas we get:

(0) Fix x_0; set $w_0 = 1$, $j = 0$ and $i = 1$.

(1) Simulate U_i; if $U_i = 0$ goto (3).

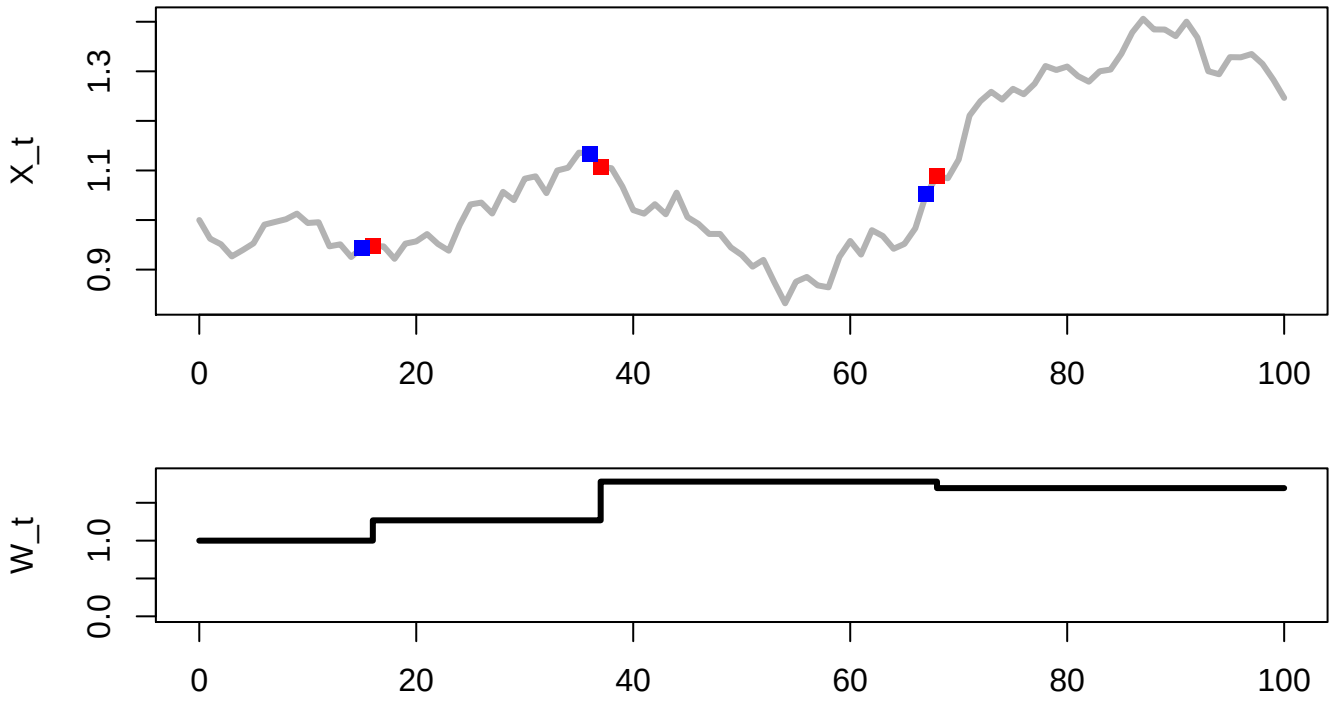
(2) $[U_i = 1]$ Simulate X_{ih} from $q(x_{ih}|x_{jh}, (i-j)h)$ and $X_{(i-1)h}$ from $q(x_{(i-1)h}|x_{jh}, x_{ih})$. Set

$$w_i = w_j r(x_{ih}, x_{(i-1)h}, h).$$

(3) Let $i = i + 1$ and goto (1).

If we stop the SIS at a time point t, then X_t can be drawn from $q(x_t|x_{jh}, t - jh)$; and the weight is w_j.

Example



Rao-Blackwellisation

At time ih, the incremental weight depends on x_{ih} and $x_{(i-1)h}$. Rather than simulating both we simulate x_{ih}, and use an expected incremental weight

$$\rho_h(x_{ih}, x_{jh}, (j-i)h) = \mathrm{E}\left(r(x_{ih}, X_{(i-1)h}, h) \mid x_{jh}\right),$$

with expectation with respect to the conditional distribution of $X_{(i-1)h}$ given x_{jh}, x_{ih} under the proposal:

$$\mathrm{E}\left(r(x_{ih}, X_{(i-1)h}, h) \mid x_{jh}\right) = \int r(x_{ih}, x_{(i-1)h}, h) q(x_{(i-1)h} | x_{jh}, x_{ih}) \mathrm{d}x_{(i-1)h}.$$

New SIS algorithm

Using these ideas we get:

(0) Fix x_0; set $w_0 = 1$, $j = 0$ and $i = 1$.

(1) Simulate U_i; if $U_i = 0$ goto (3).

(2) $[U_i = 1]$ Simulate X_{ih} from $q(x_{ih}|x_{jh}, (i-j)h)$ and set

$$w_i = w_j \rho_h(x_{ih}, x_{jh}, (i-j)h).$$

(3) Let $i = i + 1$ and goto (1).

If we stop the SIS at a time point t, then X_t can be drawn from $q(x_t|x_{jh}, t - jh)$; and the weight is w_j.

Continuous-time SIS

The previous algorithm cannot be implemented as we do not know $p(\cdot|\cdot, h)$. However, if we consider $h \to 0$ we obtain a continuous-time algorithm that can be implemented.

The Bernoulli process converges to a Poisson-process.

In the limit as $h \to 0$, if we fix $t = ih$ and $s = jh$ we get

$$\rho(x_t, x_s, t-s) = \lim_{h\to 0} \rho_h(x_t, x_s, t-s) = 1 + \frac{1}{\lambda}\left(\frac{(\mathcal{K}_x - \mathcal{K}_x^{(1)})q(x|x_s, t-s)}{q(x|x_s, t-s)}\right)\Bigg|_{x=x_t}.$$

CIS Algorithm

(0) Fix x_0; set $w_0 = 1$ and $s = 0$.

(1) Simulate the time t of the next event after s in a Poisson process of rate λ.

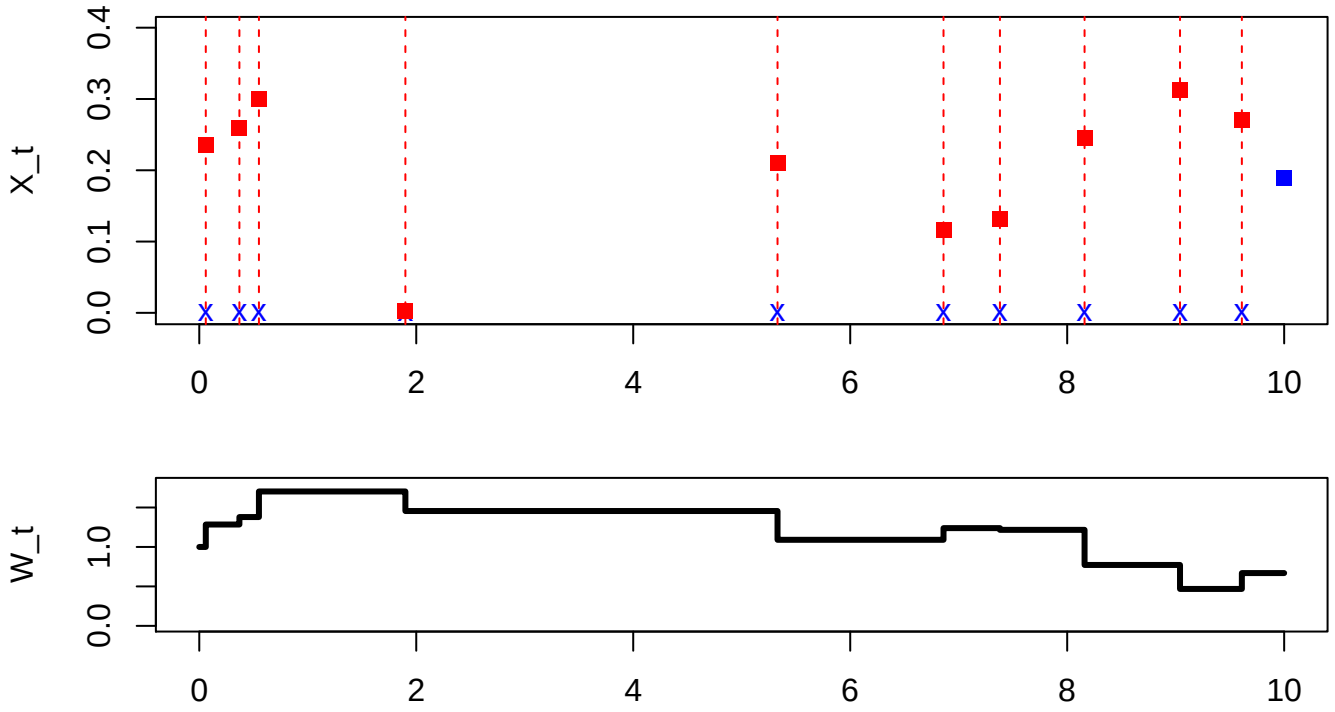
(2) Simulate X_t from $q(x_t|x_s, t-s)$; and set

$$w_t = w_s \times \rho(x_t, x_s, t-s).$$

(3) Goto (1).

If we stop the SIS at a time point T, then X_T can be drawn from $q(x_T|x_s, T-s)$; and the weight is w_j.

Example CIS



CIS for diffusions

The target process is

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t.$$

- Define an exogenous renewal process $\{\tau_1, \tau_2 \ldots\}$ with inter-arrival rate $\lambda = \lambda(t - \tau(t))$.
- Update weights at each renewal according to above formula.
- At each renewal, update the importance process:

$$\mathrm{d}X_t = b(\tau_i)\mathrm{d}t + v(X_{\tau_i})\mathrm{d}B_t.$$

Does it work?

Not always! A necessary (and it turns out sufficient) condition for the method to be valid (ie unbiased) is that the weight process $\{w_s;\ s \geq 0\}$ is a martingale. Then the CIS algorithm provides unbiased estimates of the diffusion marginal distributions (and by iterations its FDDs).

In almost all cases where the proposal is not chosen to have $v(\tau_i) = \sigma(X_{\tau_i})$ then the weight process turns out to NOT be in L^1!

What about the copycat scheme? $v(\tau_i) = \sigma(X_{\tau_i})$, $b(\tau_i) = \mu(X_{\tau_i})$

Does it work?

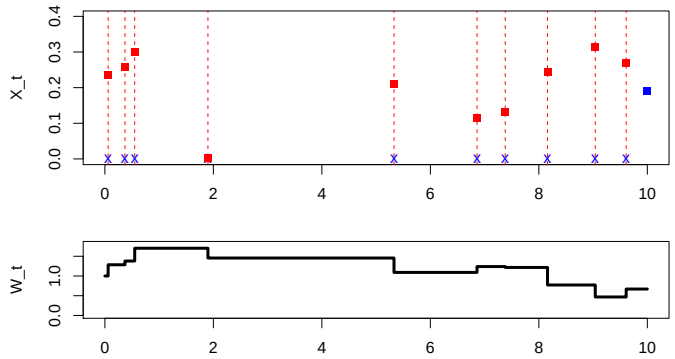
Not always! A necessary (and it turns out sufficient) condition for the method to be valid (ie unbiased) is that the weight process $\{w_s;\ s \geq 0\}$ is a martingale. Then the CIS algorithm provides unbiased estimates of the diffusion marginal distributions (and by iterations its FDDs).

In almost all cases where the proposal is not chosen to have $v(\tau_i) = \sigma(X_{\tau_i})$ then the weight process turns out to NOT be in L^1!

What about the copycat scheme? $v(\tau_i) = \sigma(X_{\tau_i})$, $b(\tau_i) = \mu(X_{\tau_i})$

Theorem:

1. If σ and μ are globally Libschitz, and σ is bounded away from 0, then the copycat scheme is valid.
2. If σ and μ are also bounded above, then for all $p > 1$, there exists $\epsilon > 0$ such that chosing $\lambda(u) \propto u^{-1+\epsilon}$ ensures that $\{w_s, s \geq 0\}$ is an L^p martingale.



$$w_T = \prod_{i=1}^{N_T} \rho_i$$

where

$$\rho_i = 1 + \frac{1}{\lambda} \left(\frac{(\mathcal{K}_x - \mathcal{K}_x^{(1)}) q(x|x_s, t-s)}{q(x|x_s, t-s)} \right) \Bigg|_{x=x_t}.$$

Comments and Extensions

For general diffusions care is needed to ensure these conditions are satisfied – we have results which give rules for implementing the procedure in these cases.

Comments and Extensions

For general diffusions care is needed to ensure these conditions are satisfied – we have results which give rules for implementing the procedure in these cases.

There is substantial extra flexibility – such as letting the Poisson rate depend on the time since the last event, or coupling the Poisson rate with the proposal process.

Comments and Extensions

For general diffusions care is needed to ensure these conditions are satisfied – we have results which give rules for implementing the procedure in these cases.

There is substantial extra flexibility – such as letting the Poisson rate depend on the time since the last event, or coupling the Poisson rate with the proposal process.

There are numerous variance reduction methods that can be used (antithetic sampling, and extra importance sampling and different proposal distribution for the process at event times).

Comments and Extensions

For general diffusions care is needed to ensure these conditions are satisfied – we have results which give rules for implementing the procedure in these cases.

There is substantial extra flexibility – such as letting the Poisson rate depend on the time since the last event, or coupling the Poisson rate with the proposal process.

There are numerous variance reduction methods that can be used (antithetic sampling, and extra importance sampling and different proposal distribution for the process at event times).

Dealing with the negative weights is an important issue.

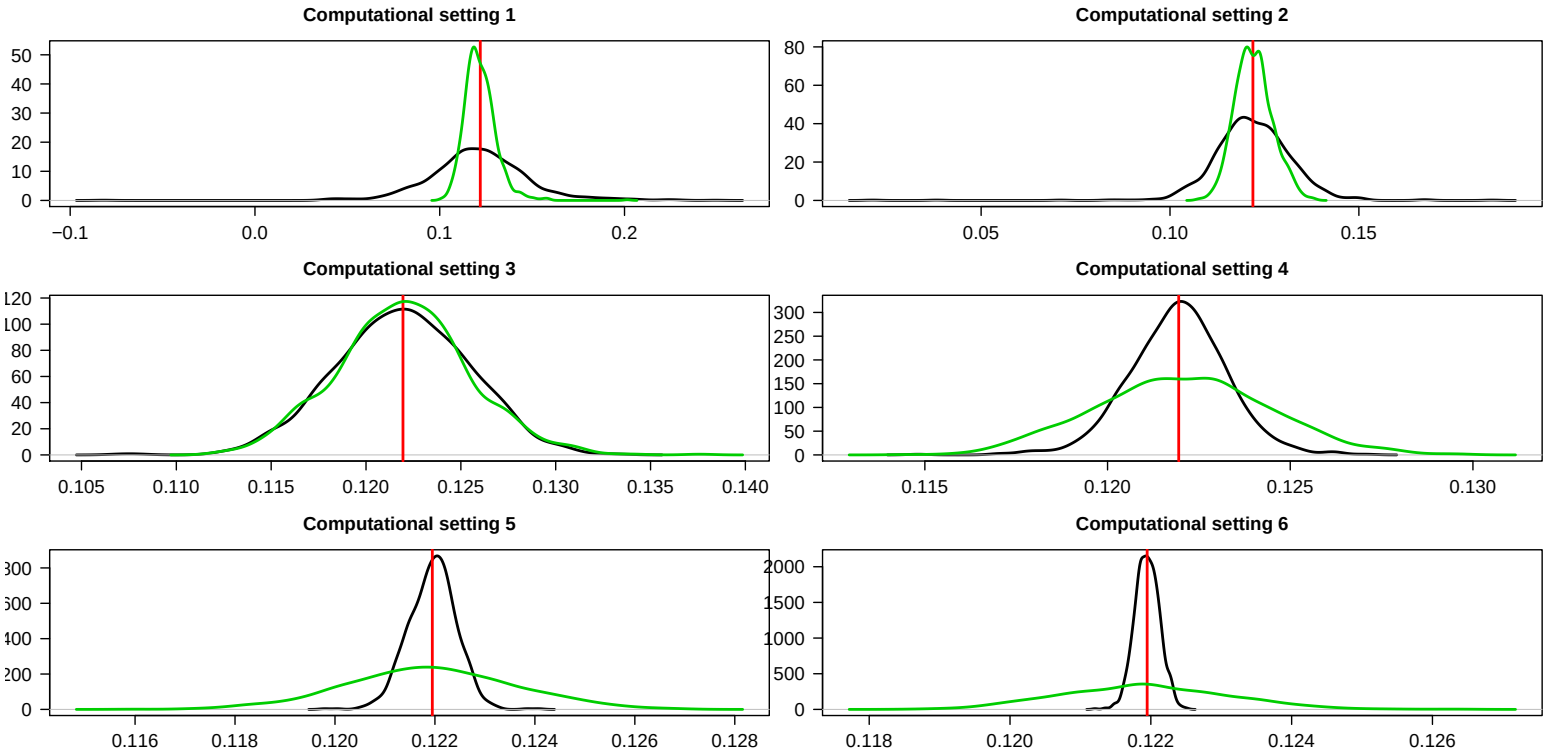
Example: CIR Diffusion

We consider estimating the transition density for a 2-d CIR model:

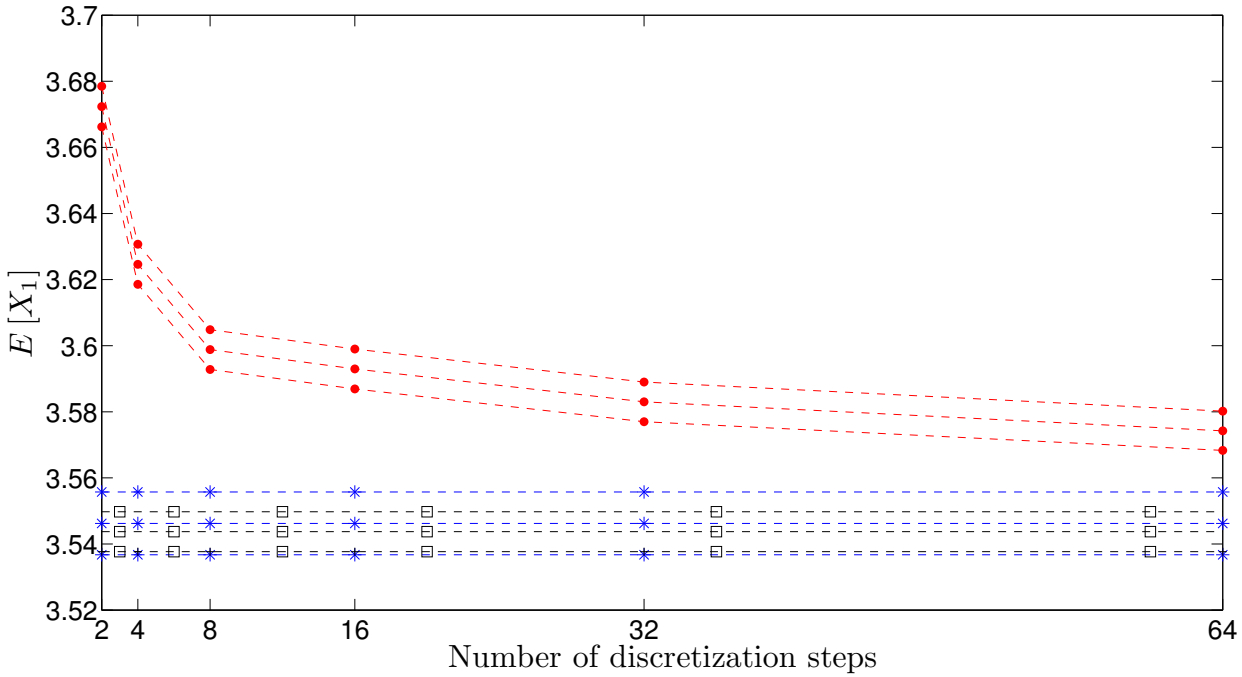
$$\begin{bmatrix} \mathrm{d}X_t^{(1)} \\ \mathrm{d}X_t^{(2)} \end{bmatrix} = \begin{bmatrix} -\rho_1(X_t^{(1)} - \mu_1) \\ -\rho_2(X_t^{(2)} - \mu_2) \end{bmatrix} \mathrm{d}t + \begin{bmatrix} \sigma_1\sqrt{X_t^{(1)}} & 0 \\ \rho\sigma_2\sqrt{X_t^{(2)}} & \sigma_2\sqrt{(1-\rho^2)X_t^{(2)}} \end{bmatrix} \begin{bmatrix} \mathrm{d}B_t^{(1)} \\ \mathrm{d}B_t^{(2)} \end{bmatrix}$$

We compare the CIS with a time-discretisation approach based on the ideas in Durham and Gallant (2002), for varying CPU cost.

Example: CIR Diffusion



Example: Jump diffusions



Discussion

New directions for the applications of Rao-Blackwell theorem.

Discussion

New directions for the applications of Rao-Blackwell theorem.

Exact simulation algorithm gives a random sufficient statistic for the drift of a diffusion process.

Discussion

New directions for the applications of Rao-Blackwell theorem.

Exact simulation algorithm gives a random sufficient statistic for the drift of a diffusion process.

Exact simulation of diffusions is possible for pretty much all one-dimensional diffusions and some multi-dimensional ones, but it is intrinsically limited in the multi-dimensional case.

Discussion

New directions for the applications of Rao-Blackwell theorem.

Exact simulation algorithm gives a random sufficient statistic for the drift of a diffusion process.

Exact simulation of diffusions is possible for pretty much all one-dimensional diffusions and some multi-dimensional ones, but it is intrinsically limited in the multi-dimensional case.

CIS is a very flexible and potentially more general method for general stochastic processes.

Discussion

New directions for the applications of Rao-Blackwell theorem.

Exact simulation algorithm gives a random sufficient statistic for the drift of a diffusion process.

Exact simulation of diffusions is possible for pretty much all one-dimensional diffusions and some multi-dimensional ones, but it is intrinsically limited in the multi-dimensional case.

CIS is a very flexible and potentially more general method for general stochastic processes.

Can be used to unbiasedly estimate density (likelihood), expectations, etc

Discussion

New directions for the applications of Rao-Blackwell theorem.

Exact simulation algorithm gives a random sufficient statistic for the drift of a diffusion process.

Exact simulation of diffusions is possible for pretty much all one-dimensional diffusions and some multi-dimensional ones, but it is intrinsically limited in the multi-dimensional case.

CIS is a very flexible and potentially more general method for general stochastic processes.

Can be used to unbiasedly estimate density (likelihood), expectations, etc

Theory established for diffusions, and very recently for jump diffusions