Exact Simulation and on-line inference for diffusions

Gareth Roberts
Lancaster University

Joint work with Alex Beskos, Paul Fearnhead and Omiros Papaspiliopoulos

Munich, October, 2006

What I wanted to do in this talk...

1. Exact simulation of diffusions
2. 'Exact' Monte Carlo likelihood based inference for diffusions
3. 'Exact' filtering for partially observed diffusions

Just attempt to overview the first and third of these.

Diffusions

Continuous, strong Markov processes described by stochastic differential equation:

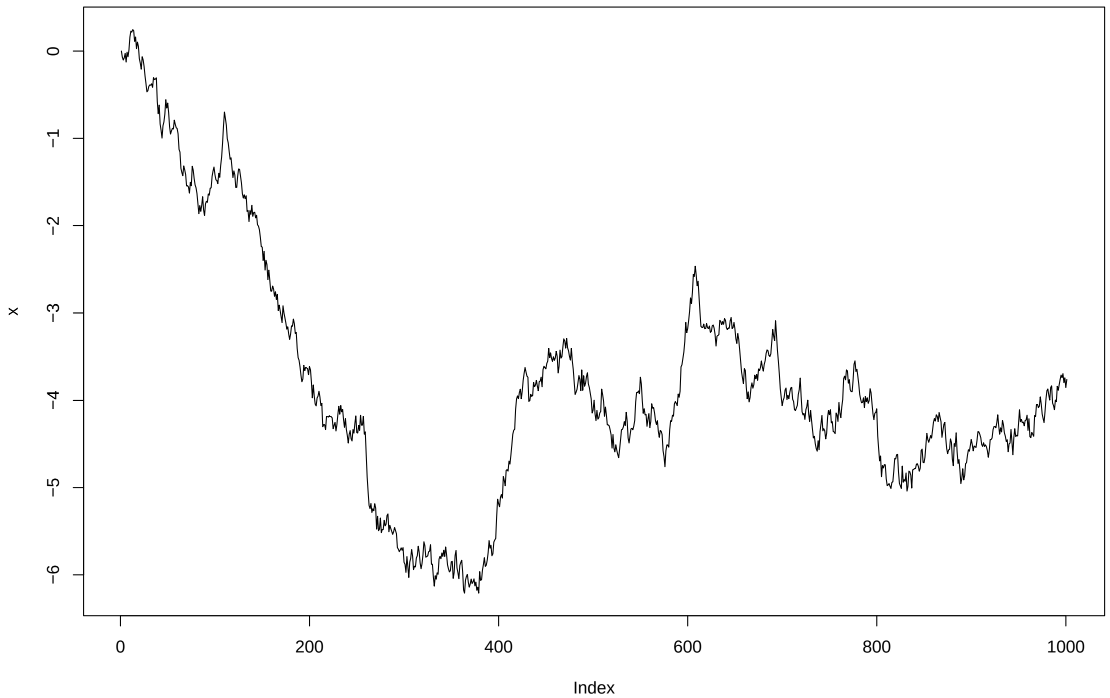
$$dX_t = \alpha(X_t)dt + \sigma(X_t)dB_t$$

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.

Motivation for simulation

- Monte Carlo estimation
- Monte Carlo ML estimation
- Imputation schemes for Bayesian inference from diffusion models
- "Exact particle filtering" for diffusion models

Diffusion densities

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

and let the law of this diffusion on $[0,1]$ be denoted $\mathbf{P}$, with $\mathbf{P}_0$ being that of the diffusion

$$dX_t = \sigma(X_t)dB_t \ .$$

Then under suitable regularity conditions

$$\frac{d\mathbf{P}}{d\mathbf{P}_0}(X) = G(X)$$

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

$$\log G(X) = \int_0^1 \left(\frac{\alpha(X_s)}{\sigma^2(X_s)} dX_s - \frac{\alpha^2(X_s)}{2\sigma^2(X_s)} ds \right)$$

Aim: to use rejection sampling...

Simulate from Brownian path

Accept with probability proportional to Girsanov evaluated at that path.

Simple!

Can this be implemented?

Assume initially that we are capable of simulating, storing, and computing integrals with whole Brownian sample paths and perform integrals numerically (with no approximation). Then we could just do **rejection sampling**.

1. Find K such that $G \leq K$ almost surely.
2. Sample a Brownian motion B.
3. Compute $G(B)$.
4. Produce an indicator variable, I, such that $P(I = 1) = G(B)/K$.
5. If $I = 0$ go to 2.
6. Output B.

The outputed sample path has the required distribution.

Problems with this rejection idea

$G(B)$ is typically unbounded.

Step 2 is impossible to perform exactly since it requires an infinite number of computations and an infinite storage capacity.

Step 3 is impossible in general at least exactly, since it requires the exact evaluation of analytically intractible integrals.

There is still hope since it may be possible to do step 4. without computing $G(B)$ explicitly. However first we need to deal with the boundeness of $G(\cdot)$ issue.

Towards a simulation algorithm: simplifying G

To fix ideas consider

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

We can rewrite

$$\log G = A(X_1) - A(X_0) - r\int_0^1 \phi(X_s)ds + \ell$$

where $\phi(X_s) = \left[\left(\alpha'(X_s) + \alpha^2(X_s)\right)/2 - \ell\right]/r$ where ℓ and r are chosen to constrain ϕ to always be in the interval $[0,1]$.

For this we need the bound

$$\ell \leq \frac{\alpha'(x) + \alpha^2(x)}{2} \leq \ell + r$$

for all $x \in \mathbf{R}$.

Firstly we consider 'biased Brownian bridge' candidates:

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

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

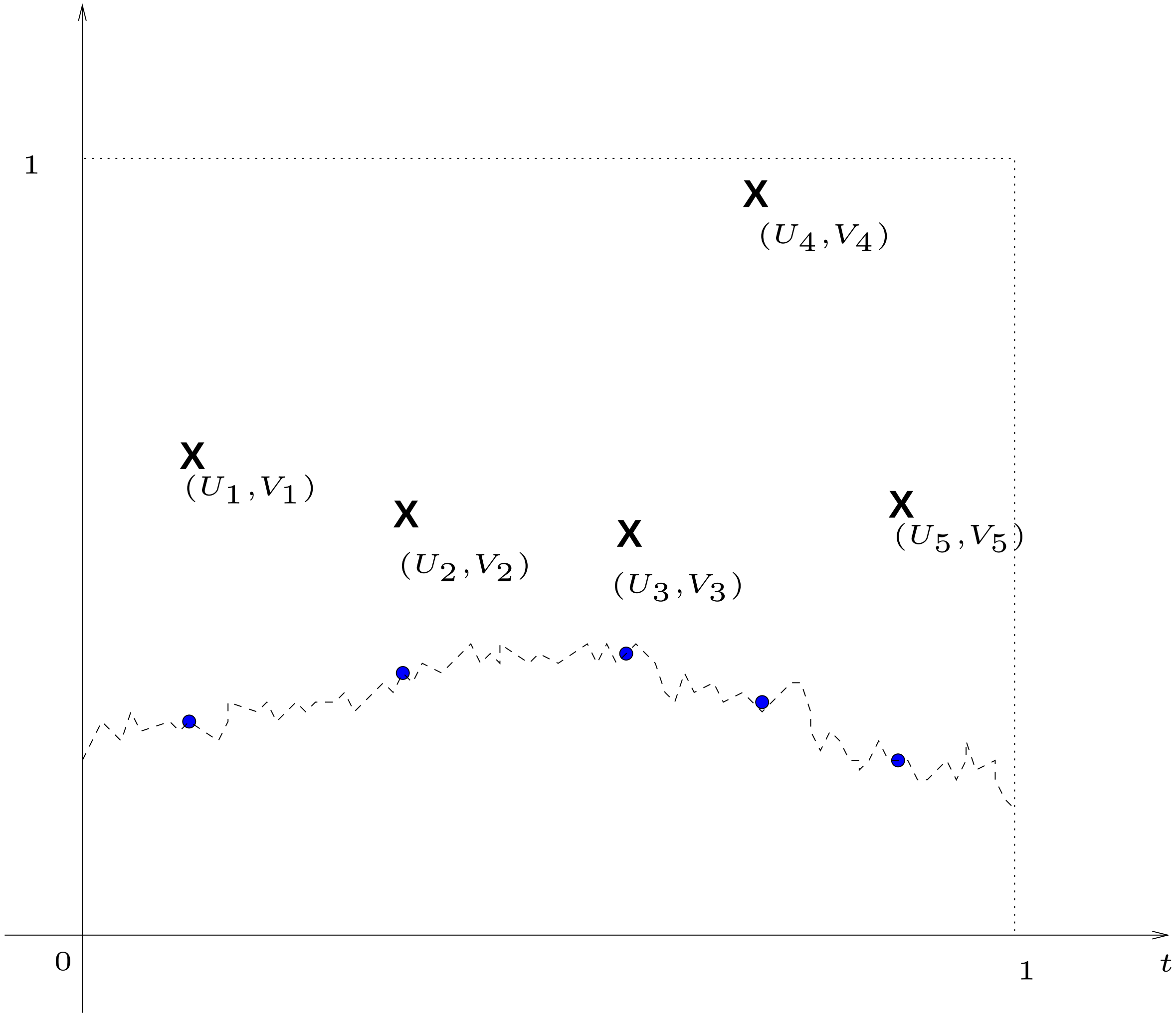
$$\frac{d\mathbf{P}}{d\tilde{\mathbf{W}}} \propto \exp\left\{-r\int_{s=0}^{1} \phi(X_s)ds\right\}.$$

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

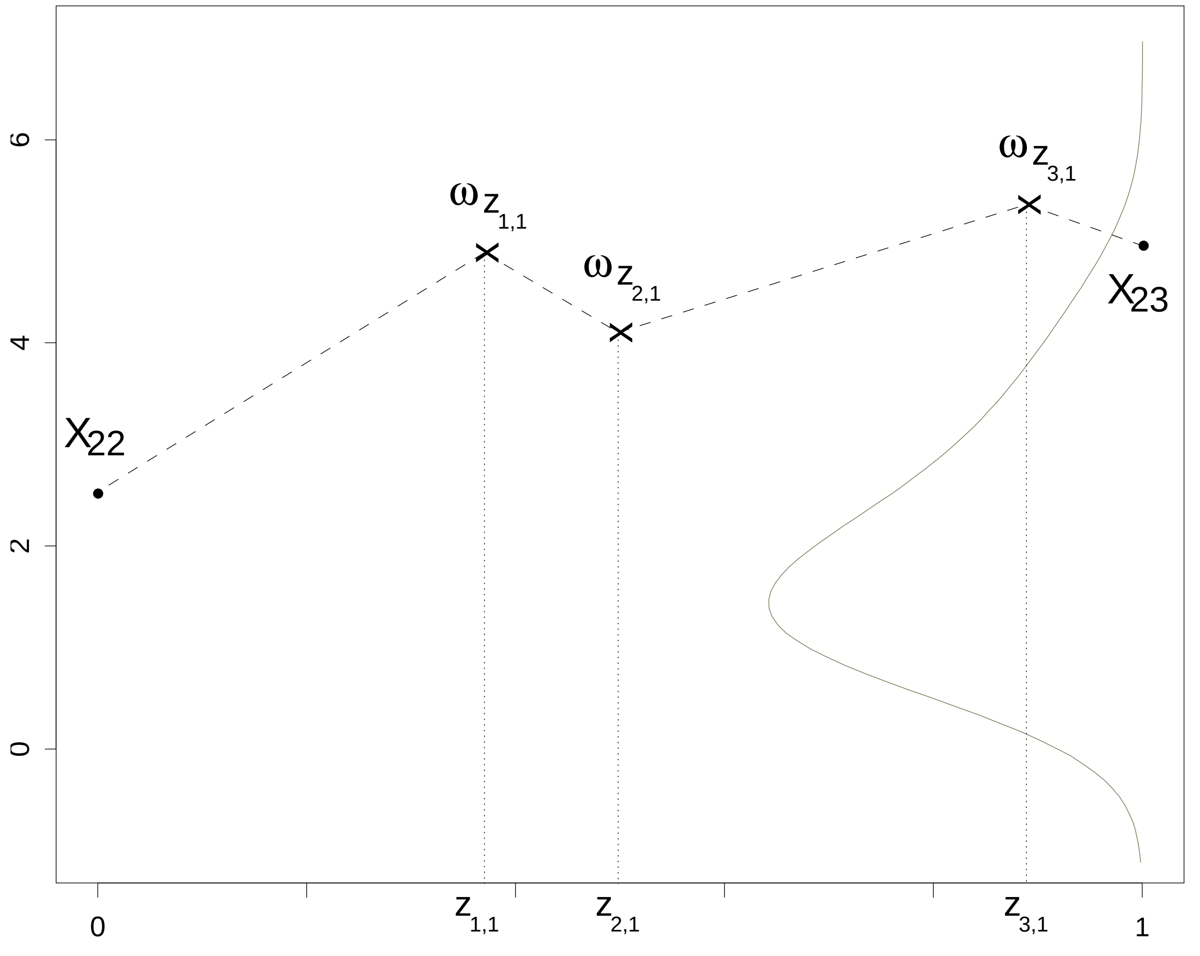
$$P(\Phi \text{ is the empty configeration} = \exp\{-r\int_0^1 \phi(X_s)ds\}).$$

The basic algorithm (EA1)

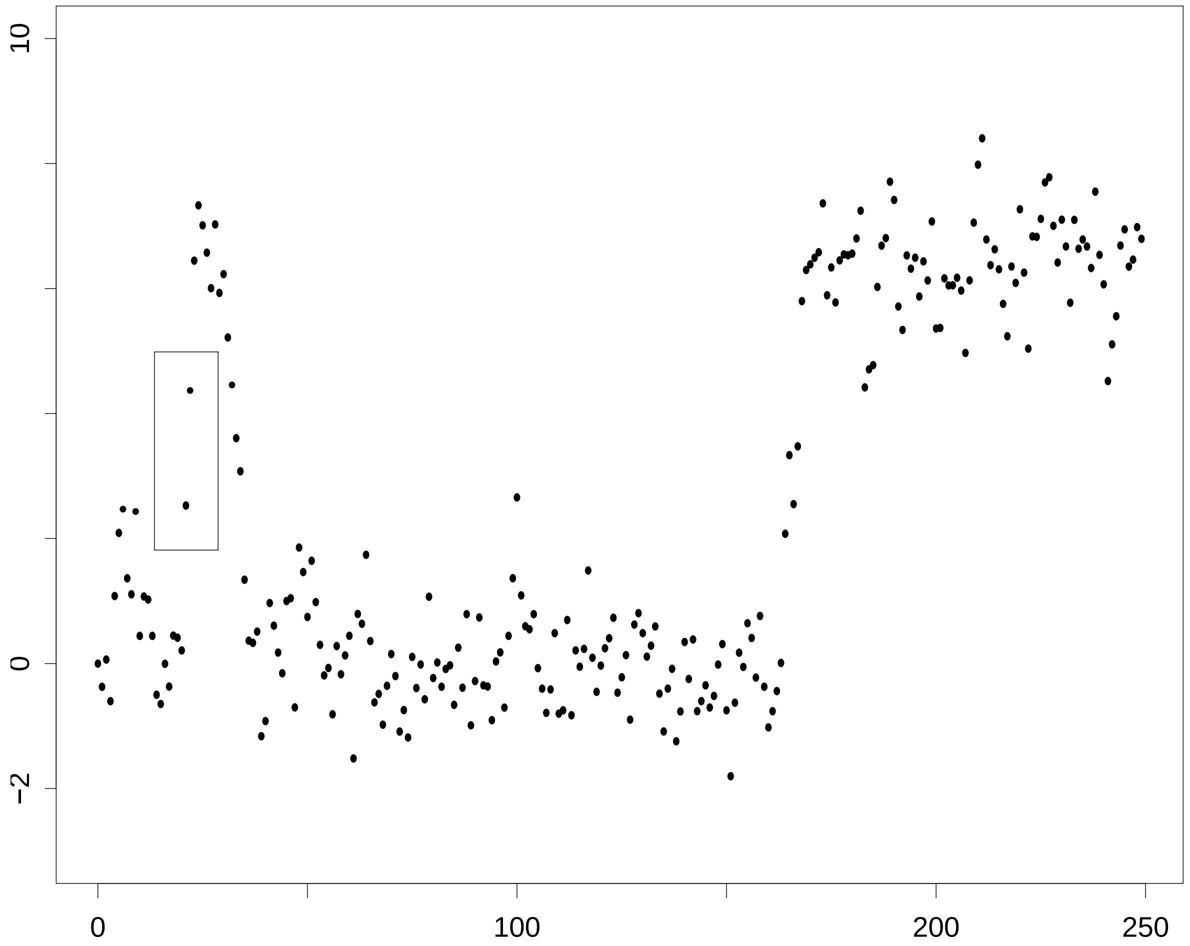
1. `Set` $B_0 = 0$. `Simulate` B_1 `from (*)`
2. `Generate Poisson process of rate` r `on` $[0,1] \times [0,1]$:
   $\Phi = \{(U_1, V_1), \ldots (U_n, V_n)\}$
3. `For each` U_i, `draw` B_{U_i} `from its appropriate Brownian bridge probabilities.`
4. `If` $\phi(B_{U_i}) > V_i$ `for ANY` i, `erase skeleton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



1
X
(U_4, V_4)
X
(U_1, V_1)
X
(U_2, V_2)
X
(U_3, V_3)
X
(U_5, V_5)
0
1
t



6
4
2
0
X_{22}
$\omega_{z_{1,1}}$
$\omega_{z_{2,1}}$
$\omega_{z_{3,1}}$
X_{23}
0
$z_{1,1}$
$z_{2,1}$
$z_{3,1}$
1


10
0
−2
0
100
200
250

How to use output

We output skeleton $\{(U_i, X_{U_i})\}$.

The skeleton can be completed at any required time t, say by merely simulating from appropriate Brownian bridge probabilities.

So suppose that U_l is the largest $U_i \leq t$ and U_u is the smallest $U_i \geq t$, we can simulate X_t from

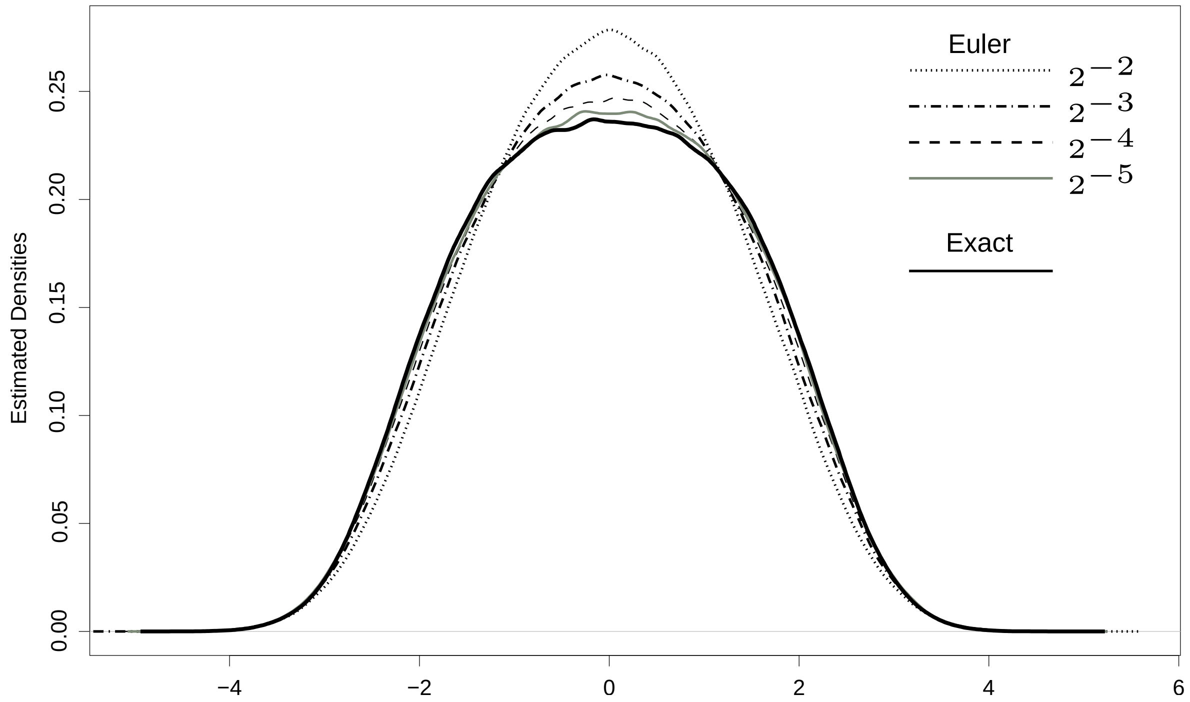
$$N\left(\frac{(t-U_l)X_{U_u} + (U_u - t)X_{U_l}}{U_u - U_l}, \frac{(t-U_l)(U_u - t)}{U_u - U_l}\right)$$

X_t is an EXACT draw from the diffusion at time t (except for finite computational precision considerations)

Simple example

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

This diffusion is not analytically tractable.



Properties of simulation algorithm

1. Method is “exact”.
2. Methods are computationally very simple.
3. Methods are computationally efficient. (Eg EA1 is linear in time interval and in the range parameter r.)

Extensions

1. EA2 and EA3 allow simulation from a very general class of diffusions, including essentially all non-explosive one-dimensional diffusions. These methods are also very efficient, though more complex.

2. Multivariate extensions are possible, though the pure simulation algorithm cannot handle the most general form for diffusion matrices and drift functions.

3. Time-inhomogeneous and jump-diffusion extensions are straightforward.

4. Simulation of diffusion bridges is simpler than unconditioned diffusions. This is crucial for the use of methodology for inference.

5. Can use EA to construct Malliavin derivatives too. There are many applications of this in finance (see work with Bruno Casella).

On-line estimation of hidden diffusions

In particle filtering problems, we can make use of the ability to simulate from the hidden diffusion component.

However it transpires that a more efficient importance sampling scheme which exploits the same framework, is more efficient.

System Model

We consider inference for a d-dimensional diffusion satisfying

$$\mathrm{d}X_t = \nabla A(X_t)\mathrm{d}t + \mathrm{d}B_t$$

given partial observations at discrete time-points $t_1, t_2, \ldots, t_n$.

The transition density can be written as (Beskos et al. 2006; Dacunha-Castelle, Florens-Zmirou, 1986)

$$p(x_t|x_0) = \mathcal{N}(x_t - x_0; t)\exp\{A(x_t) - A(x_0)\}\mathbb{E}\left(\exp\left\{-\int_0^t \phi(w_s)\mathrm{d}s\right\}\right),$$

where $\phi(x) = (\nabla A^T \nabla A + \nabla^2 A)/2$, and the expectation is with respect to a Brownian bridge between x_0 and x_t.

Notice: tractable/intractable terms, change of measure

Observation Regimes

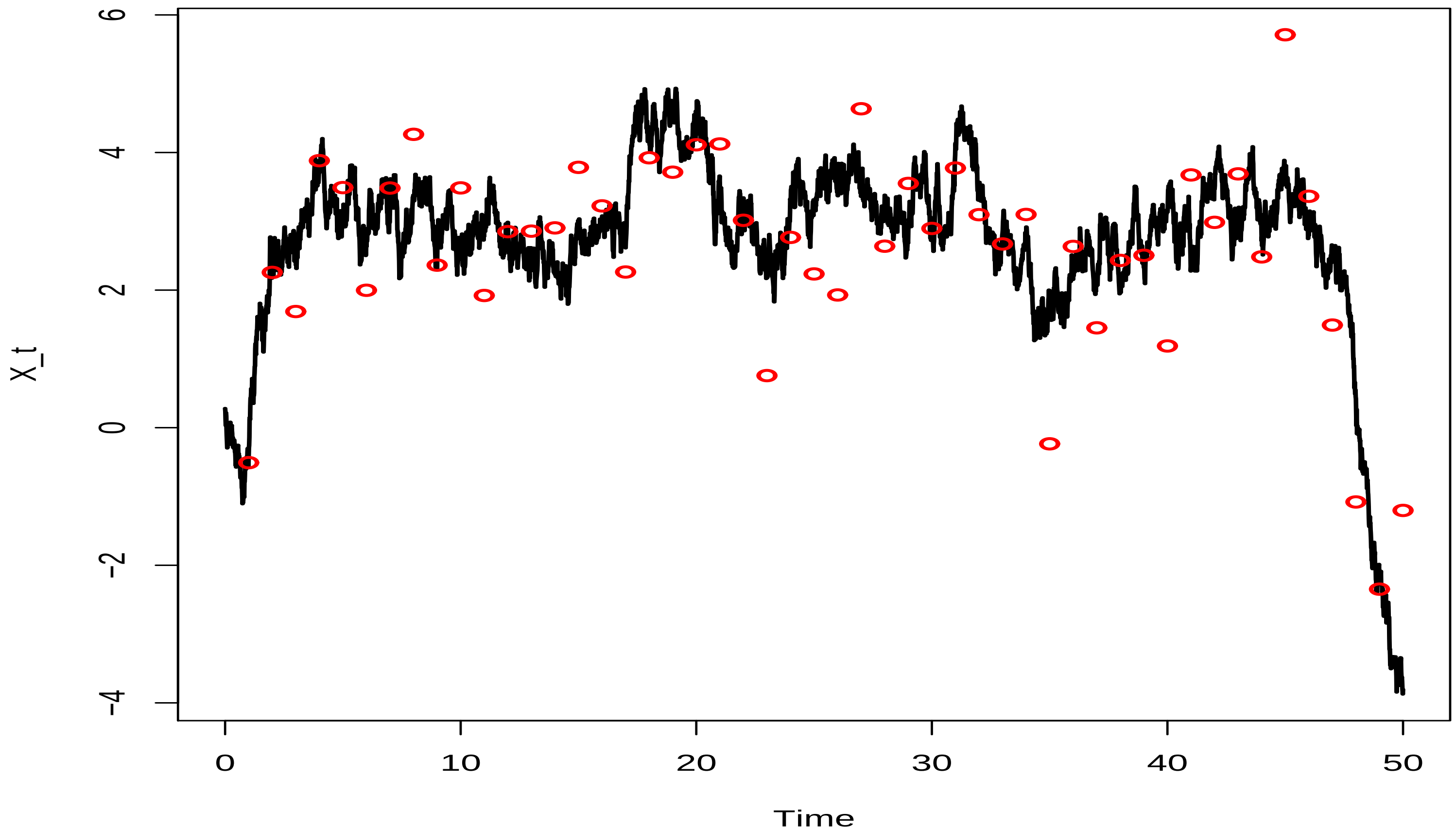
(A) Diffusion observed with error. The observation at time t_i, y_i, is related to the state of diffusion at time t_i via a known likelihood function $f(y_i|x_{t_i})$.

(B) Partial Information. At time t_i we observe $Y_i = g(X_{t_i})$ for some non-invertible known function $g(\cdot)$.

(C) Cox Process. We observe the events of a Poisson process of rate $\lambda(X_t)$.

Focus on (A) and assume $t_i = i$.

SINE data



$$dX_t = \sin(X_t)dt + dB_t, \qquad Y_i \sim N(X_i, \sigma^2).$$

Simple but interesting: non-stationary, "meta-stable" signal, intractable transition density, $X \bmod 2\pi$ is stationary.

Particle Filter

Particle filters approximate $p(x_i|y_{1:i})$ by a set of N particles, $\{x_i^{(j)}\}_{j=1}^N$, and associated weights $\{w_i^{(j)}\}_{j=1}^N$.

This gives an approximation to $p(x_{i+1}|y_{1:i+1})$,

$$\tilde{\pi}_{i+1}(x_{i+1}) \propto \sum_{j=1}^{N} w_i^{(j)} p(y_{i+1}|x_i^{(j)}, x_{i+1}) p(x_{i+1}|x_i^{(j)}).$$

We aim to get a particle approximation to this density.

Auxillary Particle Filter

The ASIR filter of Pitt and Shephard gives a generic approach to obtaining this particle approximation.

We choose a current particle $x_i^{(j)}$ with some probability β_j; we then simulate a new particle from some distribution $q(x_{i+1}|x_i^{(j)})$.

The new particle x_{i+1} is given weight

$$\frac{w_i^{(j)} p(y_{i+1}|x_i^{(j)}, x_{i+1}) p(x_{i+1}|x_i^{(j)})}{\beta_j q(x_{i+1}|x_i^{(j)})}.$$

Random Weight Particle Filter

The problem with applying this ASIR filter is that the weights are intractable due to the

$$\mathbb{E}\left(\exp\left\{-\int_i^{i+1}\phi(w_s)\mathrm{d}s\right\}\right)$$

term in the system transition density.

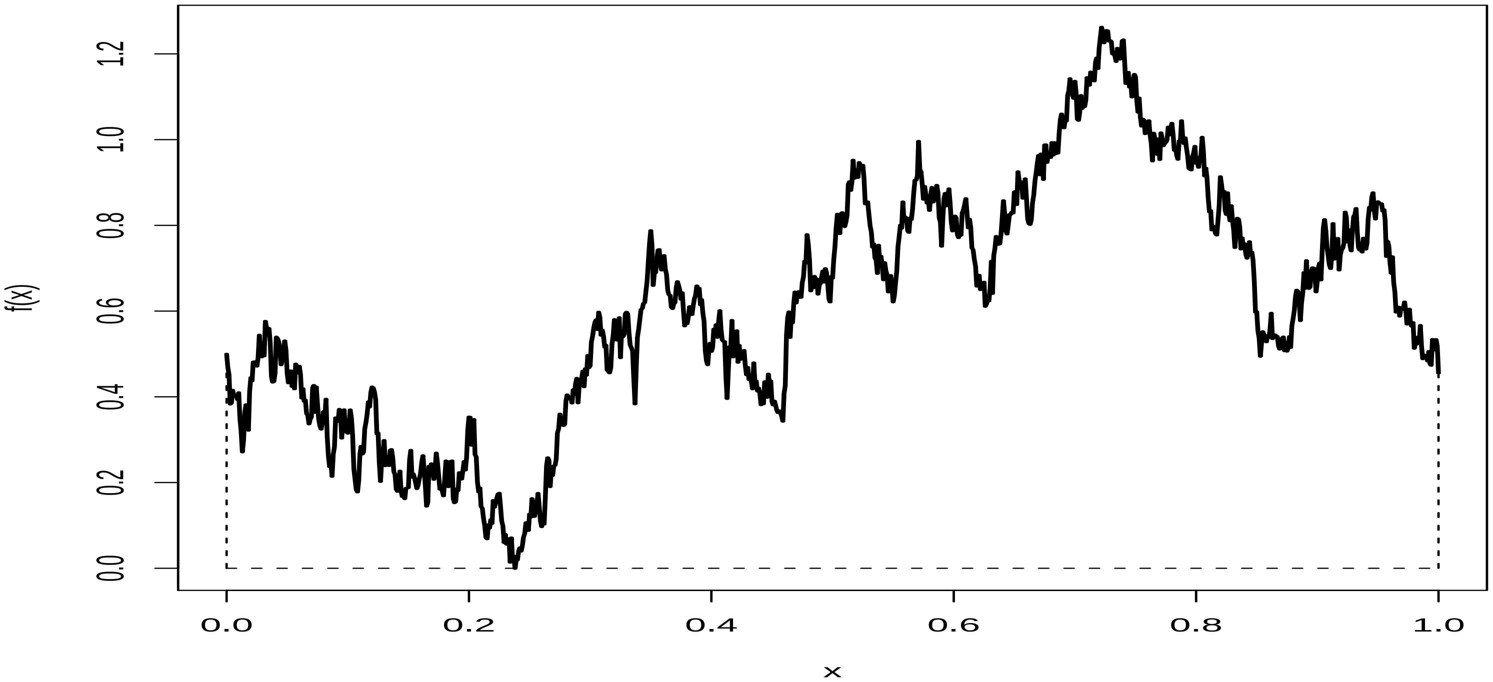
The exact algorithm simulated from events with proability proportional to this. More efficient here to do importance sampling.

Instead we introduce random weights; obtained by simulating a positive random variable whose mean is equal to the true weight.

Aside: Estimating Integrals

Consider estimating the integral

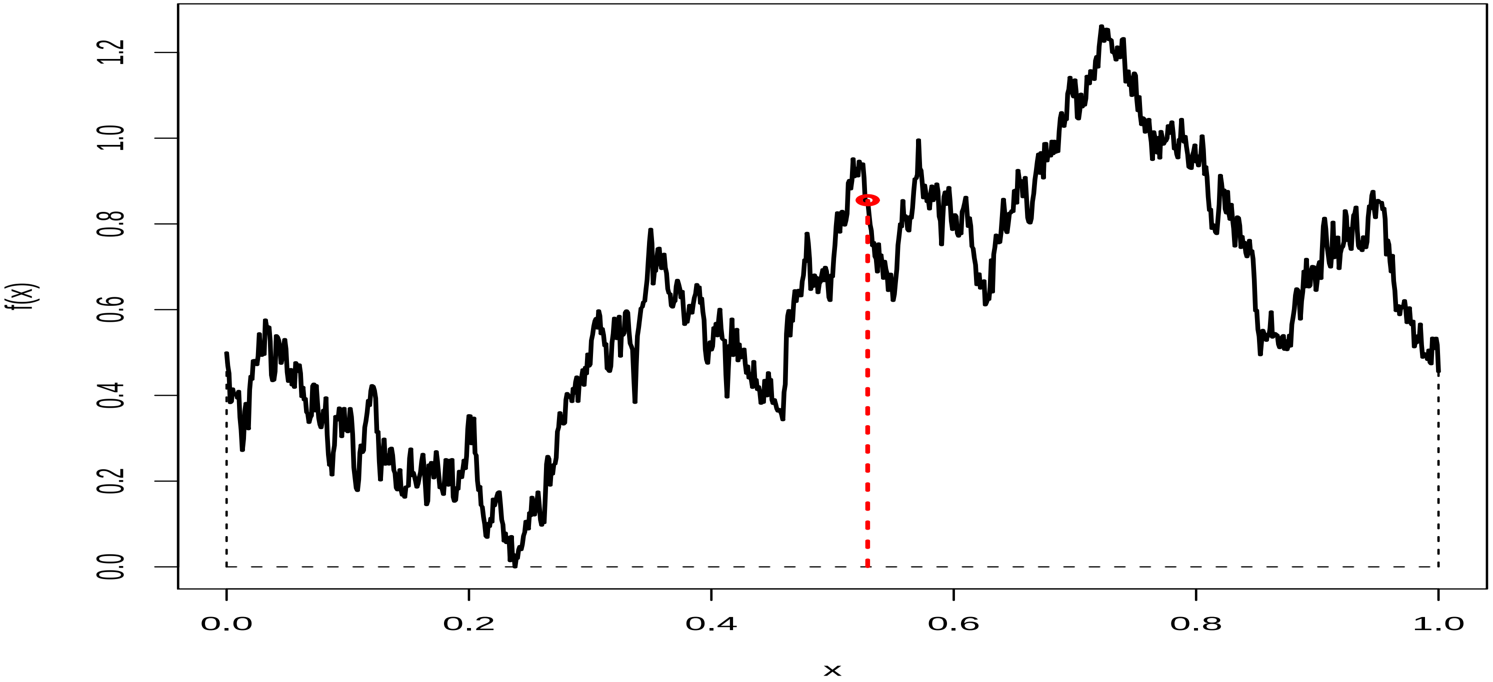
$$I = \int_0^t f(x)\mathrm{d}x.$$



Simple Unbiased Estimator

Simulate $X_1 \sim \mathrm{U}[0, t]$ and

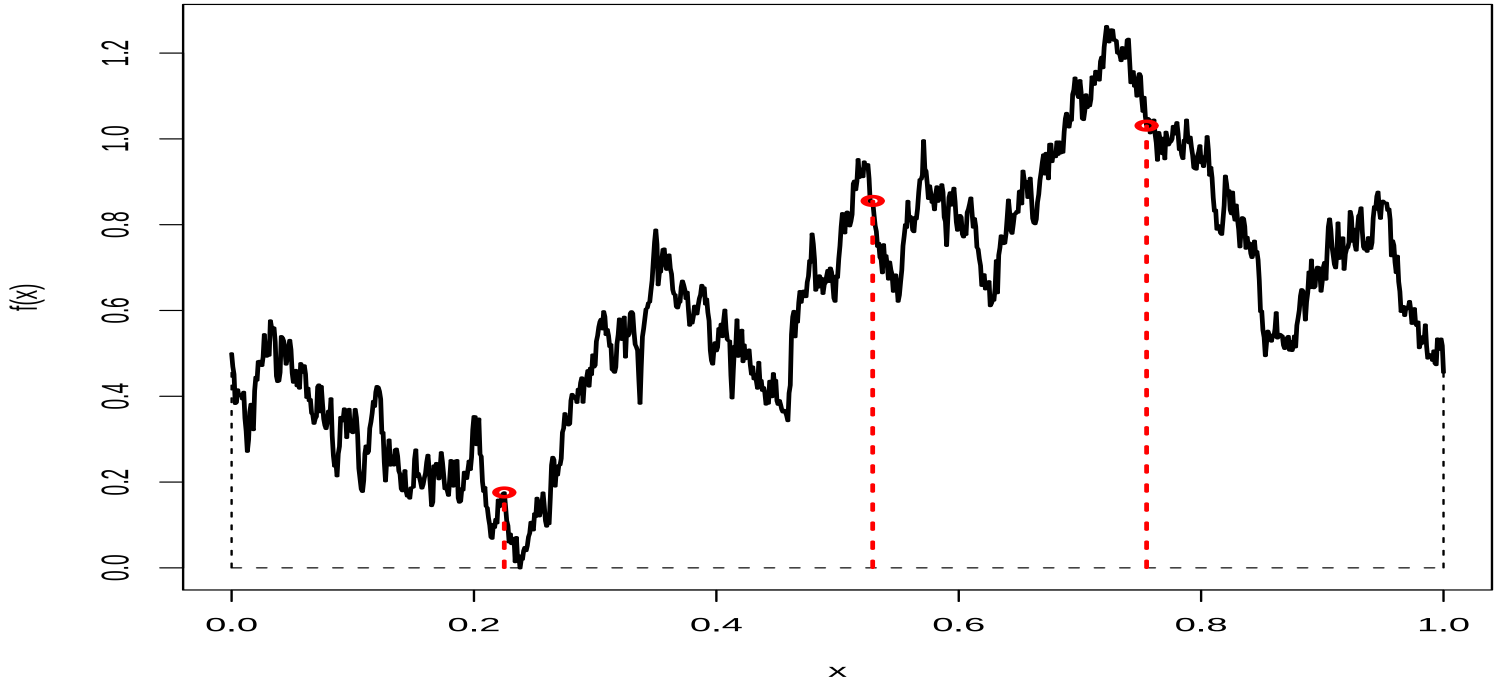
$$\hat{I} = tf(x_1).$$



Unbiased Estimator (2)

An estimate of I^k is obtained from independent $X_j \sim \mathrm{U}[0, t]$

$$\hat{I}^k = \prod_{j=1}^{k} \left(t f(x_j) \right).$$



Generalised Poisson Estimator

We can write

$$\mathbb{E}\left(\exp\left\{-\int_0^t \phi(w_s)\mathrm{d}s\right\}\right) = \mathbb{E}\left(\exp\{-ct\}\sum_{k=0}^{\infty}\left\{\int_0^t [c-\phi(w_s)]\mathrm{d}s\right\}^k/k!\right)$$

There is a simple unbiased estimator of

$$\left\{\int_0^t [c-\phi(w_s)]\mathrm{d}s\right\}^k$$

based on simulating the value of a Brownian Bridge at k independent time points in $[0,t]$.

An unbiased estimator of an infinite sum $\sum_{k=0}^{\infty} a_k$ can be obtained by Importance Sampling: choose a distribution $p(k)$; simulate k; and estimator is $a_k/p(k)$.

Important questions of tuning $p(k)$s and c for efficiency....

Random Weight Filter Theory

Central Limit Theorem of Chopin (2004) applies under extra conditions on the variance of the random weights. Our result is based on an state-space-expansion which includes also the random weight.

This means a rate of convergence of $N^{-1/2}$ for CPU cost N. This is better than filters based on discretising time, as these require increasing the number of particles and having finer discretisations. These filters may have rates of convergence $N^{-1/3}$ for observations regimes (A) and (B) for CPU cost N. (See e.g. Crisan, Del Moral and Lyons 1999)

SINE example

We consider analysing the SINE example. We compare

(i) A simple filter based on using Exact Simulation of the system equation to propagate particles.

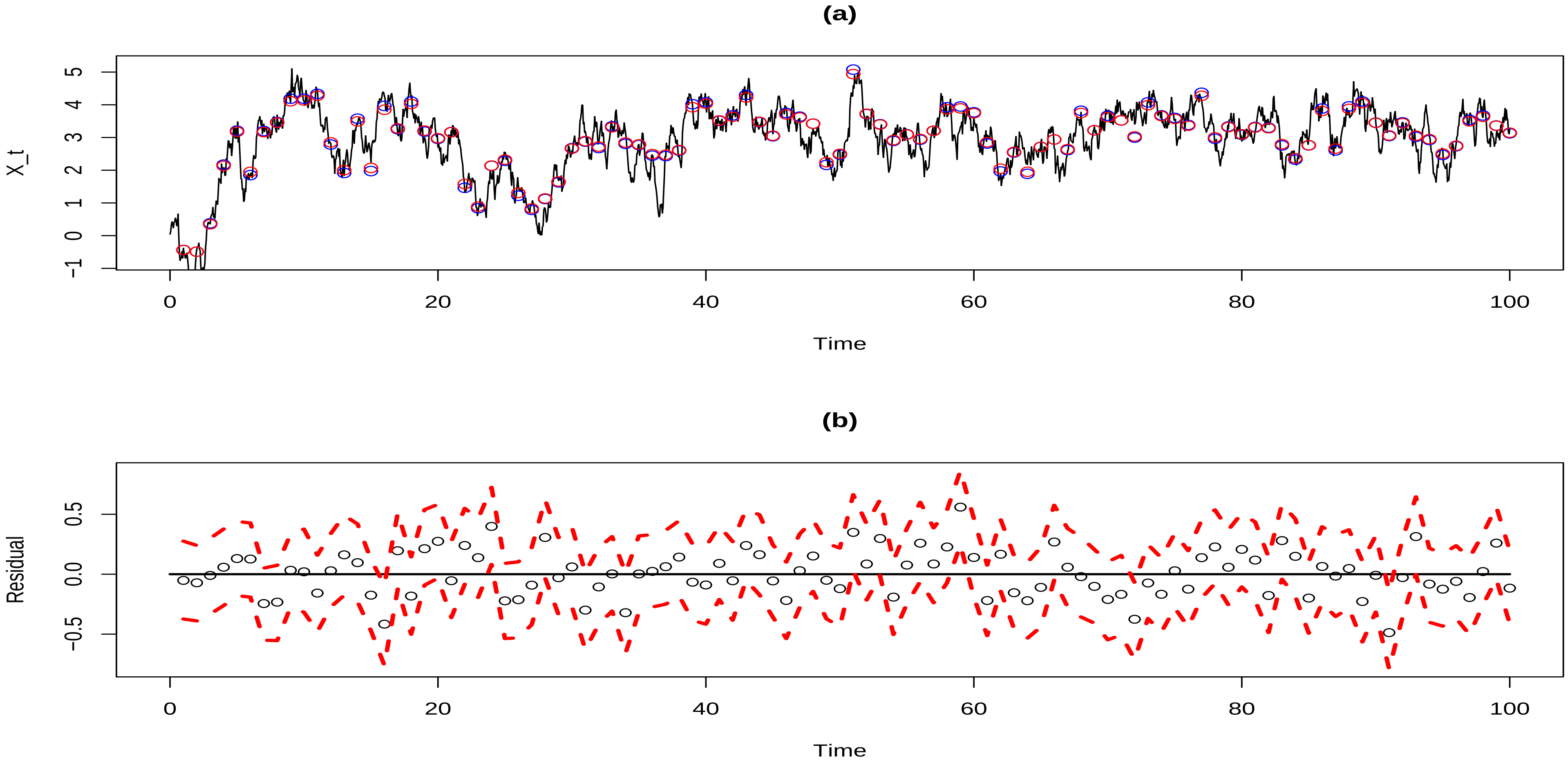
(ii) An exact rejection sampling filter - based on adapting the idea of Hurzeler and Kunsch.

(iii) The Random Weight Particle filter. Proposal distribution uses information from observation.

In each case we fixed N so that they had comparable CPU costs.

Average error of estimate of posterior mean from (iii) is reduced by 65% over (i); and 30% over (ii)

SINE data Results (2)



COX example

We use the Random Weight Particle Filter to analysis data from the following point process model:

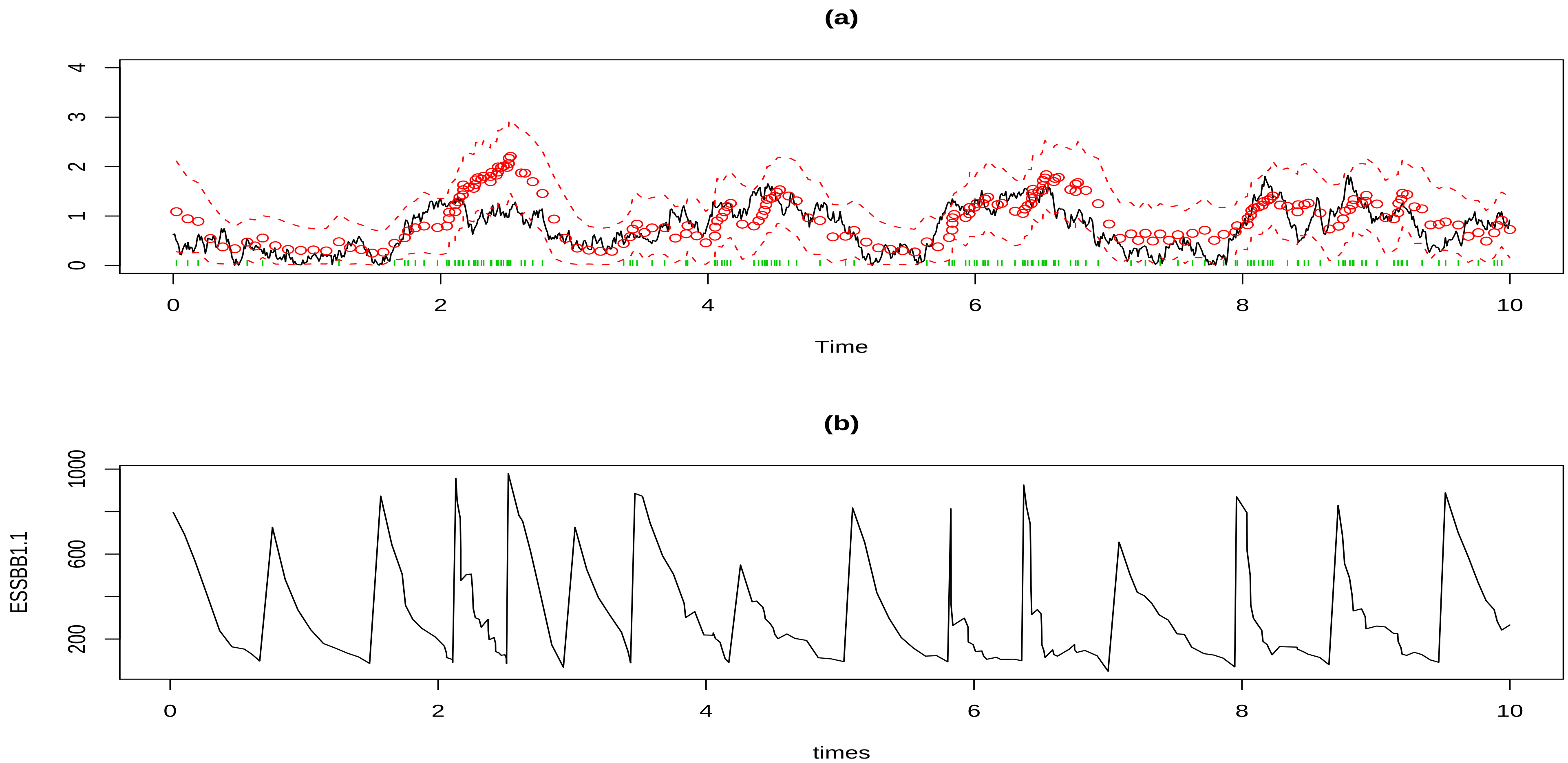
The state, X_t is an Ornstein-Uhlenbeck (OU) process,

$$\mathrm{d}X_t = -\frac{1}{2}\rho X_t \mathrm{d}t + \mathrm{d}B_t \,.$$

and we observed points of a conditional Poisson process with rate $20|X_t|$.

For this model, the likelihood of the data depends on the complete (continuous) path of the system state.

COX data Results (1)



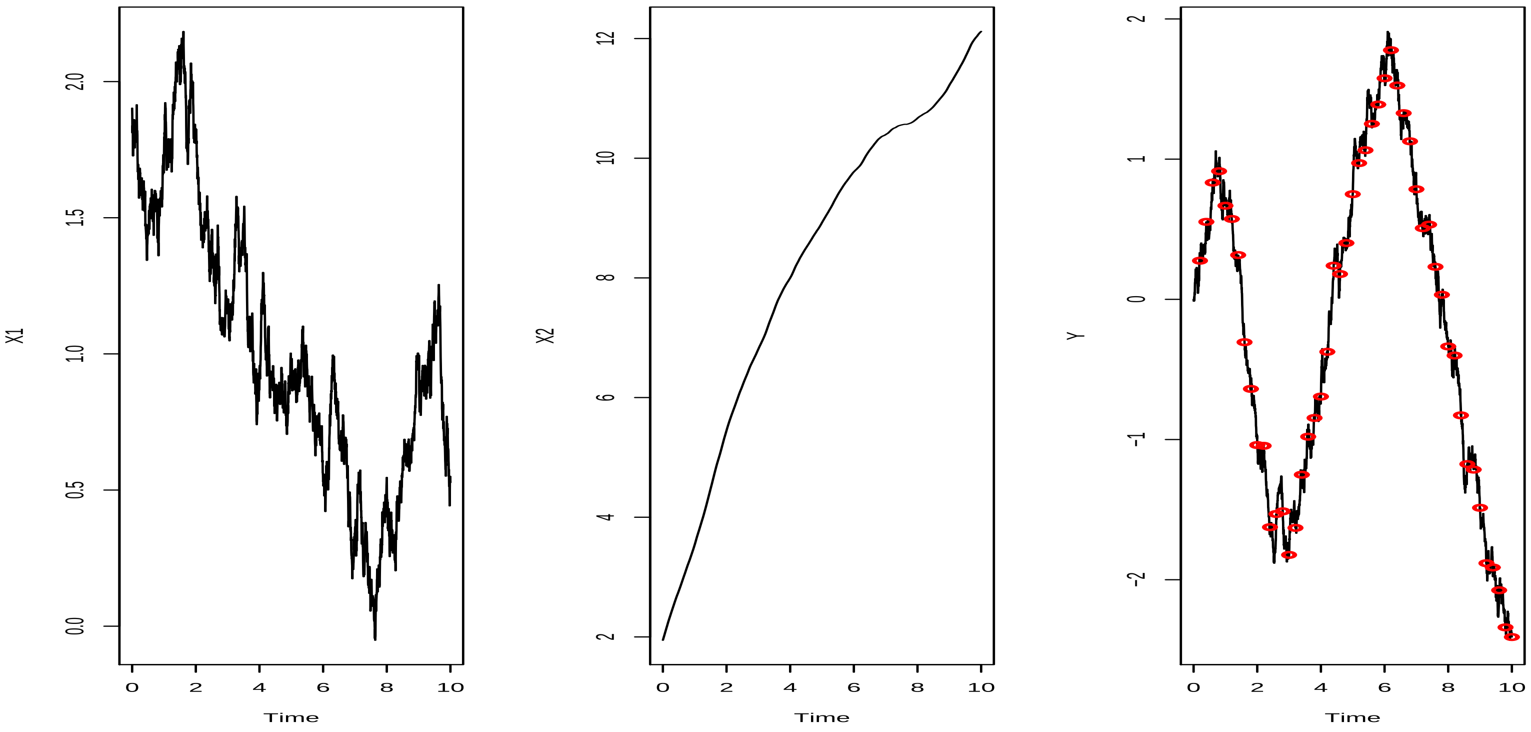
Hypo-elliptic Example

A model for FM demodulation:

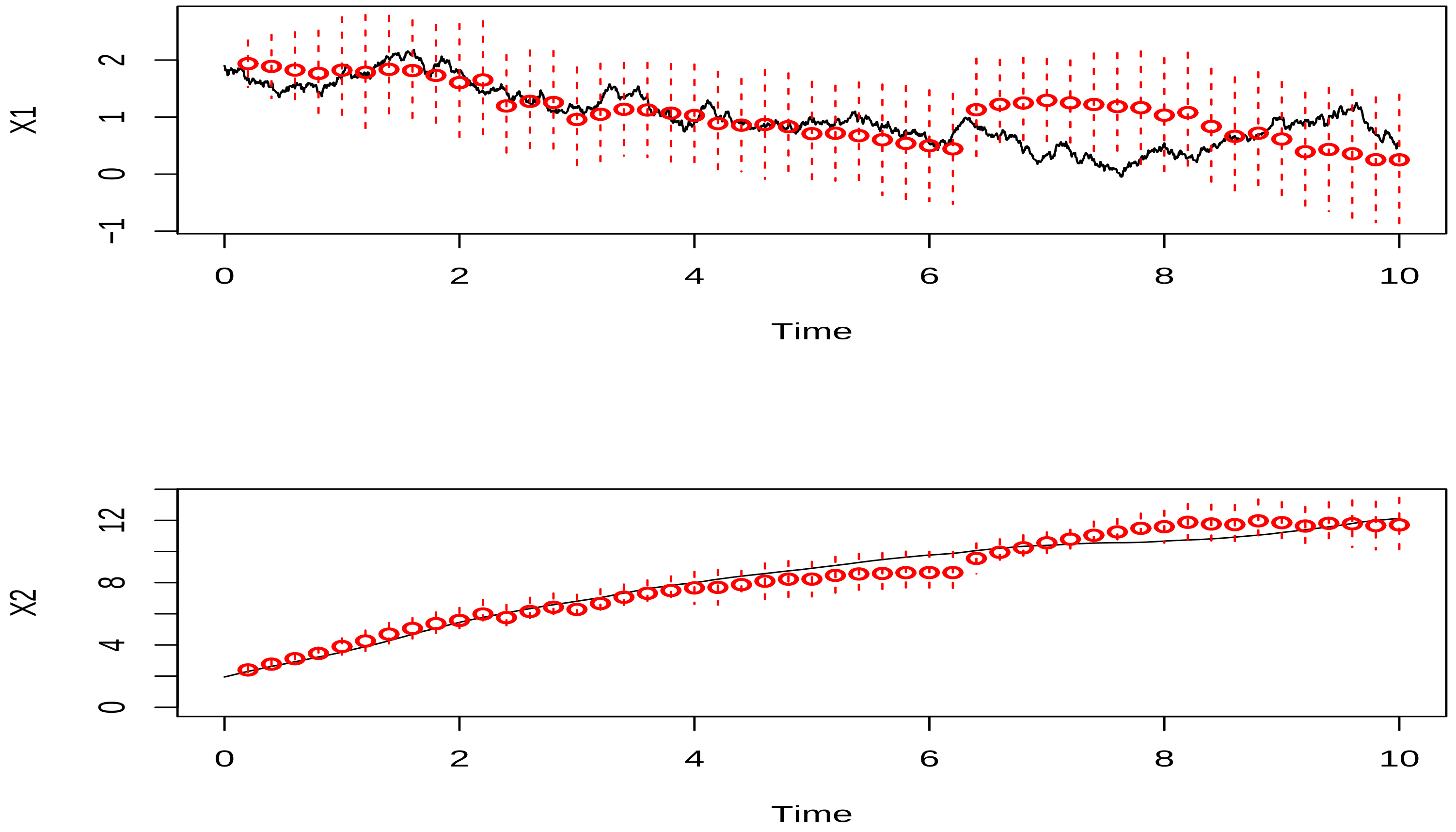
$$\begin{aligned} \mathrm{d}X_t^{(1)} &= -\alpha X_t^{(1)}\mathrm{d}t + \sqrt{2\gamma\alpha}\,\mathrm{d}B_t^{(1)} \\ \mathrm{d}X_t^{(2)} &= X_t^{(1)}\mathrm{d}t \\ \mathrm{d}Y_t &= \sqrt{2\gamma}\sin(\omega t + X_t^{(2)})\mathrm{d}t + \sigma\mathrm{d}B_t^{(2)} \end{aligned}$$

Consider inference given exact observations of Y_t at discrete time points.

Data


X1
2.0
1.5
1.0
0.5
0.0
0 2 4 6 8 10
Time
X2
12
10
8
6
4
2
0 2 4 6 8 10
Time
Y
2
1
0
-1
-2
0 2 4 6 8 10
Time

Results



Extensions

There are important open problems which require a combination of theory and application of standard/non-standard computational statistical methodology.

For example: extension to models with non-constant volatility (e.g. Stochastic Volatility); alternative approaches to ensure positivity of weights; use of variance reduction techniques in the generalised Poisson Estimator; automatic procedures for designing proposal distributions.

Considerable scope for blending together the exact simulation methodology with this general importance sampling framework.

References

See `http://www.maths.lancs.ac.uk/~papaspil/research.html`

Beskos, A. O. and Roberts, G. O. (2003) Exact Simulation of Diffusions, *Ann. Appl. Prob.*, 15, 2422 - 2444, 2005.

Beskos, A. Papaspiliopoulos, O. and Roberts, G. O. (2006) Retrospective Exact Simulation of Diffusion Sample Paths with Applications, to appear in *Bernoulli*.

Beskos, A. Papaspiliopoulos, O. and Roberts, G. O. (2005) Monte Carlo maximum likelihood estimation for discretely observed diffusion processes.

Beskos, A. Papaspiliopoulos, O. Roberts, G. O. and Fearnhead P. N. Exact and efficient likelihood–based estimation for discretely observed diffusion processes, (with discussion) *J. Roy. Stat. Soc.* , B, 68, 3, 333-382, 2006.

Fearnhead, P. N. Papaspiliopoulos, O. and Roberts, G. O. (2005) Particle Filters for partially observed diffusions.