Enabling Quantification of
EQUIP
Uncertainty for Inverse Problems

Enabling Quantification of Uncertainty for Inverse Problems.

£2.6M project largely sponsored by EPSRC.

Led by Andrew Stuart, me, Mark Girolami and Mike Christie.

`http://www2.warwick.ac.uk/ffac/sci/maths/research/grants/equip`

Bayesian analysis of Inverse problems. Particular focus on problems in Subsurface Geophysics

4 postdoc positions available soon: three at Warwick, the fourth at Heriot-Watt University. Each position is for two years with an option for extension up to four years.

Launch day: 9th October

Retrospective Simulation

Gareth Roberts
University of Warwick

Finnish Society of Sciences and Letters
Helsinki, September 2013

This talk

1. The magic of stochastic simulation
2. What is retrospective simulation?
3. Simulation of diffusions

The magic of stochastic simulation

Many problems in science involve the computation of integrals:

$$S = \int_{\mathcal{X}} f(\mathbf{x}) d\mathbf{x}$$

or (just as hard) sums:

$$S = \sum_{\mathcal{X}} f(\mathbf{x}) \ .$$

To be concrete, assume that $\mathcal{X} = \{0,1\}^d$, the set of all sequences of length d. So the size of $\mathcal{X}$ is 2^d. Also suppose $f(x) \geq 0$ for all x.

Direct enumeration of the sum is prohibitively expensive unless d is small.

The Monte Carlo Method

von Neumann (1940s)

Monte Carlo: Take $X_1, \ldots X_n$ to be randomly chosen sequences uniformly drawn from $\mathcal{X}$. Estimate

$$\hat{S} = \frac{2^d}{n} \sum_{i=1}^{n} f(X_i)$$

As $n \to \infty$, $\hat{S} \to S$.

Catch! Often need to have $n \geq O(2^d)$ for this estimator to have an acceptably small variance.

Try again: importance sampling

We want to concentrate more on regions where there is a high f value. Instead write:

$$S = \sum_{\mathcal{X}} \frac{f(\mathbf{x})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x}$$

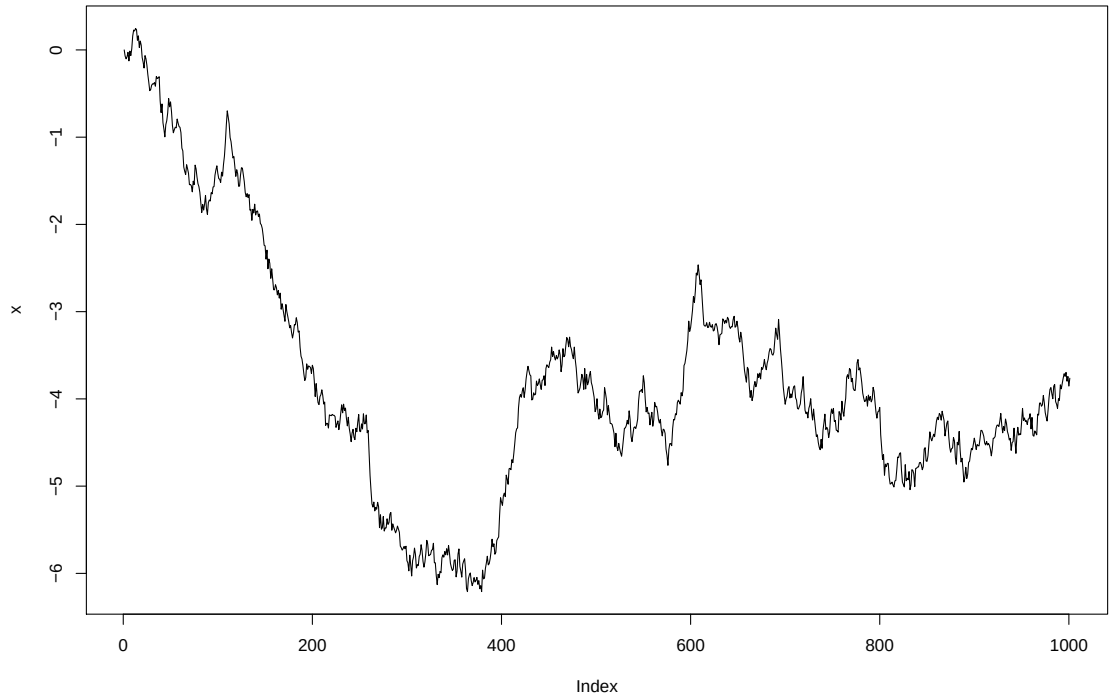
Where g is a probability distribution over $\mathcal{X}$ which matches f as closely as possible. (It might even be proportional to f.)

Catch! g is now typically difficult to simulate from.

Good ways of carrying out these simulations typically use Markov chain Monte Carlo, Sequential Monte Carlo and sometimes rejection and importance sampling.

Methods are generally less expensive than $O(2^d)$ though virtually always at least $O(d^k)$ for some $k > 0$.

What happens in infinite dimensional problems?



Traditional approaches involve approximation, truncation, discretisation to reduce to a finite-dimensional simulation problem.

What is Retrospective simulation?

Randomness and order in the exact sciences.

Retrospective simulation is an attempt to take advantage of the redundancy inherent in modern simulation algorithms (particularly MCMC, rejection sampling) by subverting the traditional order of algorithm steps.

Exact randomness by changing order.

It is (in principle) very simple!

Retrospective simulation is most powerful in infinite dimensional contexts, where its natural competitors are approximate and computationally expensive. In contrast, restrospective methods are often computationally inexpensive and "exact".

Ex 1: The birth of retrospective simulation?

Consider the quiz question on a Children's television programme:

Who is the Prime Minister of Finland?

1. Sami Hyypia
2. Jyrki Katainen
3. Sauli Niinisto

N people enter a competition to win a prize, entering their answer on a postcard. The winner is drawn uniformly from those who get the question right (ie most of them). Suppose a proportion $p > 0.5$ get it right.

Algorithm 1

1. Mark each of the N entries, placing the correct postcards into a bucket.

2. Shake the bucket and then pick out one postcard, declaring its author the winner.

Cost of this procedure, $O(N)$.

Algorithm 2

1. Throw all the postcards into the bucket without marking them

2. Draw postcards until a winner is found

Cost of this procedure, $O(p^{-1})$.

Ex. 2: The alternating series method

Devroye (1986)

Let $p = a_0 - a_1 + a_2 - a_3 + a_4 - \ldots$, where $\{a_i\}$ is a decreasing sequence. To simulate an event of probability p, the retrospective method is as follows.

Use partial sums as upper and lower bounds for p:

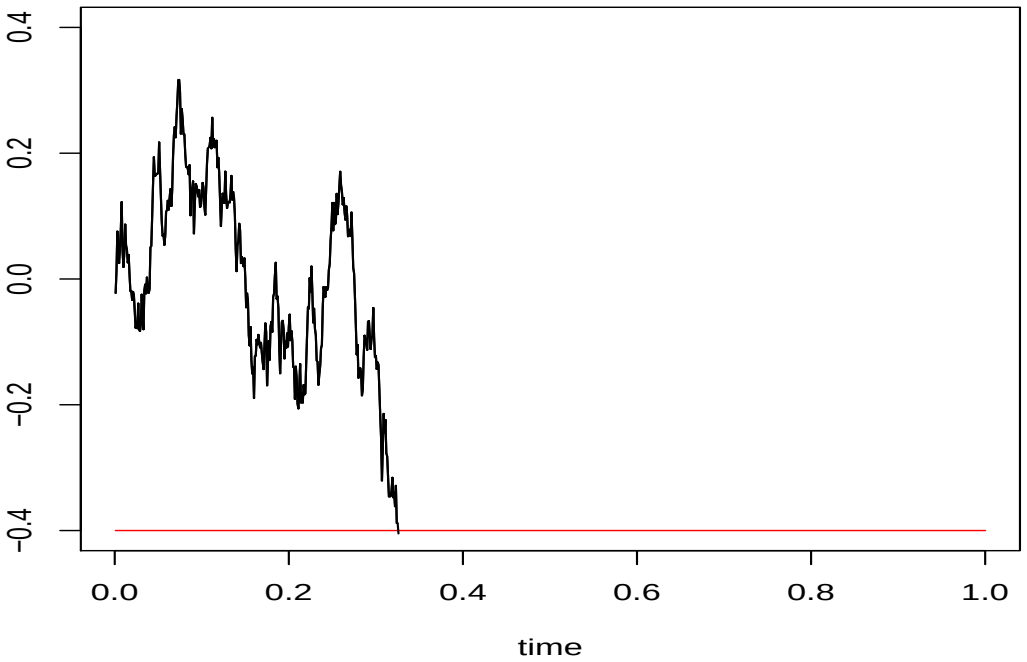
$$p_i^+ = \sum_{j=0}^{2i} a_j(-1)^j;$$

$$p_i^- = \sum_{j=0}^{2i-1} a_j(-1)^j;$$

1. Simulate $U \sim U(0,1)$.
2. Find i with both p_i^+ and p_i^- are either above or below U
3. When values are less than U, event is true, otherwise false.

Example: Simulation of BM hitting times

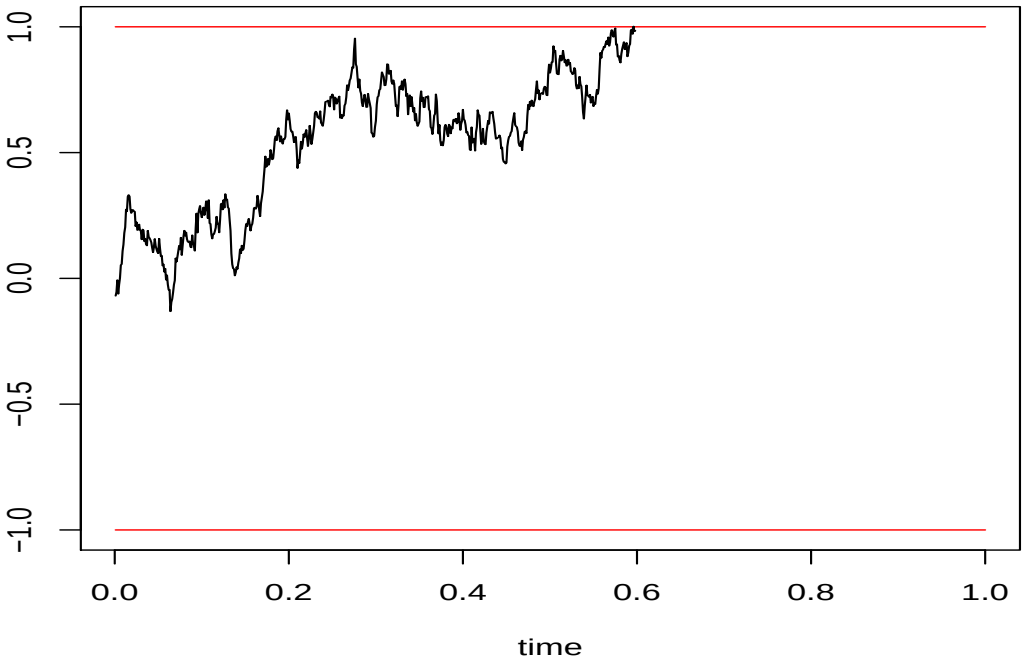
Let B_t be standard Brownian motion. Let $\tau_a = \inf\{t;\ B_t = a\}$



The distribution of τ_a is readily available analytically:

$$\mathbf{P}(\tau_a > t) = 2\Phi\left(\frac{-|a|}{t^{1/2}}\right)$$

Consider two-sided hitting time, $\tau_{a,-b} = \inf\{t;\ B_t = a \text{ or } -b\}$. Harder.



However by the reflection principle there exists an expansion

$$\mathbf{P}(\tau_{a,-b} \leq t) = a_0 - a_1 + a_2 \ldots$$

so we can apply the alternating series method.

Ex 3: Simulating from unnormalised probabilities

We have $p_1, p_2, \ldots$ is a sequence of positive numbers with $p_i \leq q_i$ and $\sum_{i=j+1}^{\infty} q_i = G(j) < \infty$.

We would like to simulate from the discrete distribution with probabilities proportional to $\{p_i\}$.

Why not use the inverse CDF method?

1. Calculate $s = \sum_{i=1}^{\infty} p_i$
2. Simulate $U \sim U(0, 1)$.
3. Set $X = \inf\{j;\ \sum_{i=1}^{j} p_j/s \geq U\}$.

Retrospective inverse CDF method

$$s_j^- = \sum_{i=1}^{j} p_i$$

$$s_j^+ = \sum_{i=1}^{j} p_i + G(j)$$

Clearly

$$s_j^- \le s_{j+1}^- \le s \le s_{j+1}^+ \le s_{j+1}^+$$

$$P_i^{+j} = \sum_{k=1}^{j} \frac{p_k}{s_j^-}$$

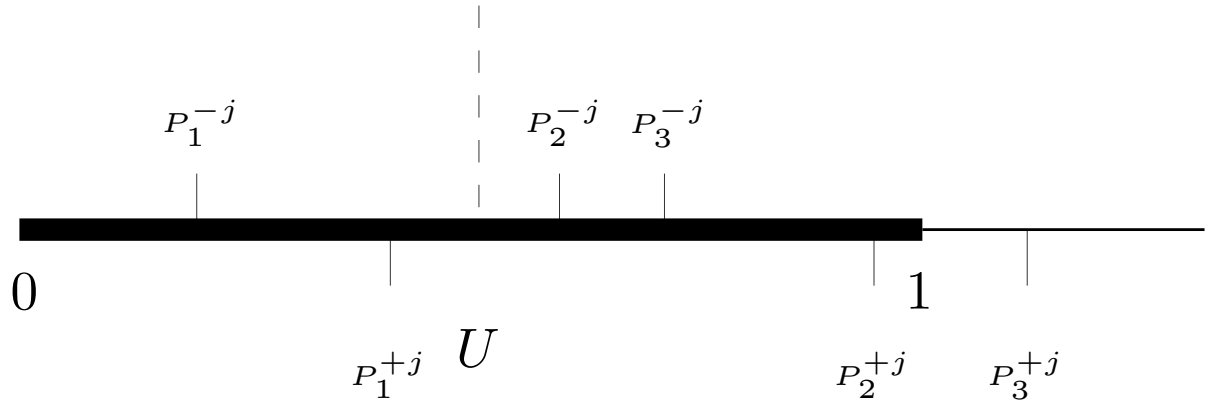
$$P_i^{-j} = \sum_{k=1}^{j} \frac{p_k}{s_j^+}$$

$$X^{+j}(U) = \inf\{j;\ P_i^{+j} \ge U\}$$

$$X^{-j}(U) = \inf\{j;\ P_i^{-j} \ge U\}$$

1. Simulate $U \sim U(0,1)$.

2. Calculate $X^{-j}(U)$ and $X^{+j}(U)$, $j = 1, 2, \ldots$ until $X^{-j}(U) = X^{+j}(U)$. Set X to be this common value.



Here $X^{+j} = X^{-j} = X = 2$.

Ex. 4: Retrospective MCMC

Many opportunities.

Peeking forward at future observations ....

Eg simulate from $\pi(\theta, X)$ with θ 'simple' and X 'complex'.

Consider Gibbs sampler which alternates between updating $\theta|X$ and $X|\theta$. The latter step is harder than the former.

However by suitable construction of random map $X \mapsto \theta$ (eg by catalytic field coupler, Breyer + R, 2001) we can often avoid having to calculate 'all' of X.

Ex. 5: Coupling from the past

Propp and Wilson (1996).

Here the naive sampler starts at time $-\infty$ from all possible states. It then records the chain value at time 0.

CFTP starts at time 0 and proceeds backwards till the chain value at time 0 is inevitable.

Ex 6: 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 probability measure $\mathbf{P}_0$.

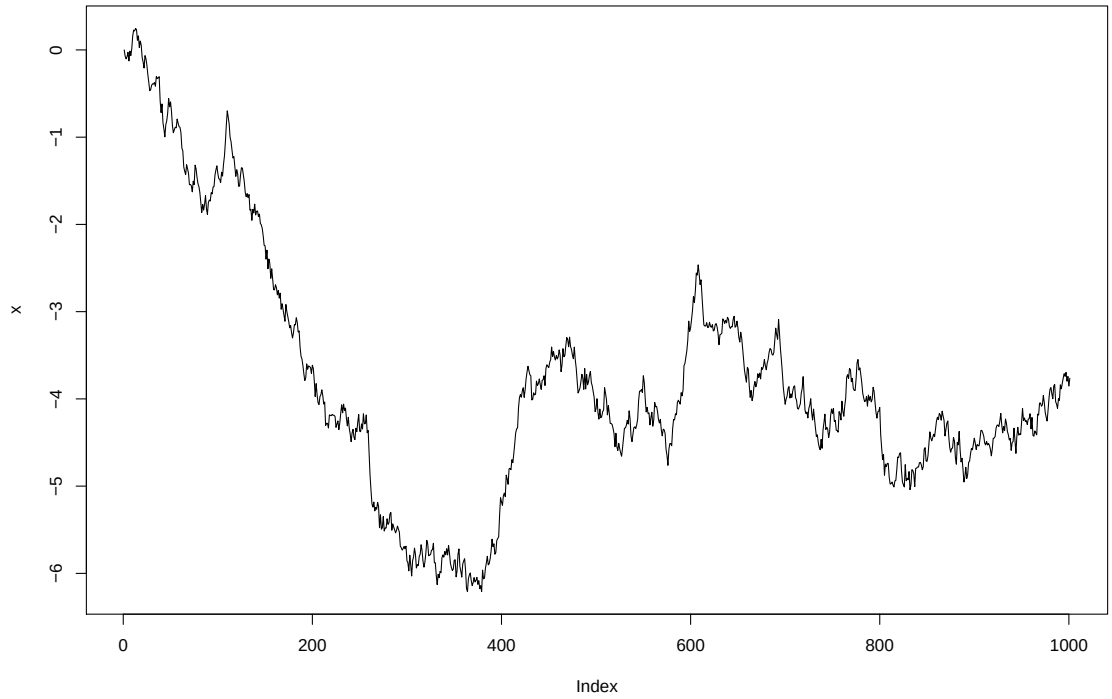
Suppose we are able to simulate from $\mathbf{P}_0$.

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

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

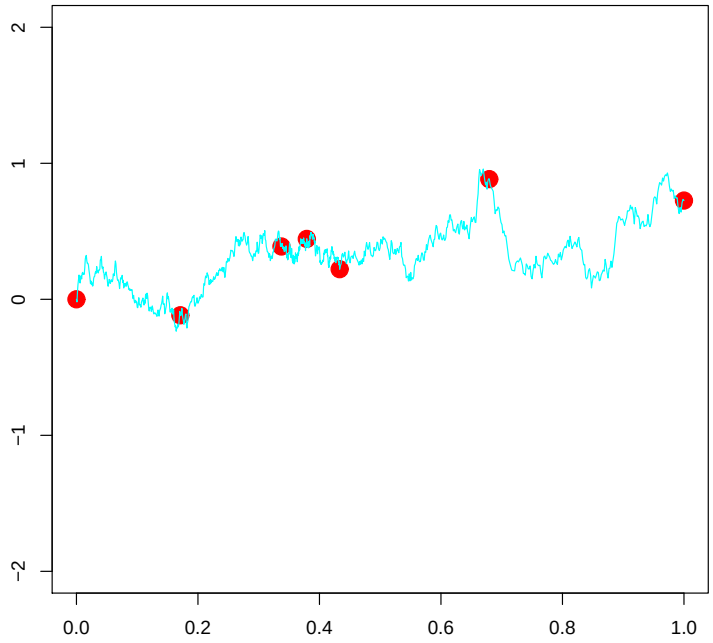
for some function ϕ taking values in $[0,1]$.

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.

The exact algorithm for diffusion processes output



Beskos, Papaspiliopoulos and R (2006, Bernoulli)

The red points act as the DNA for the entire trajectory.

Rejection for sample paths

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

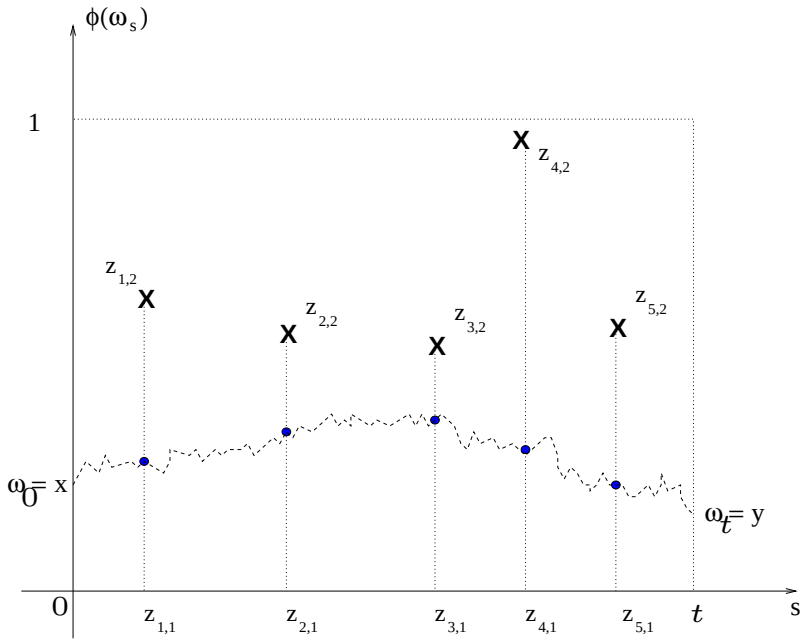
- Just storing all of X could require infinite storage capacity.
- Calculating $\int_0^1 \phi(X_s)ds$ is likely to require infinite computation

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 $\{(u, v) \in [0, 1]^2;\ v \leq \phi(u)\}$.



Slightly more general ....

Again we are able to simulate from $\mathbf{P}_0$, but want to simulate from $\mathbf{P}$:

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

If Φ is unbounded below then rejection sampling intrinsically fails.

If Φ is bounded below but is an unbounded function above, then rejection sampling still works in principle, but how can we simulate from the epigraph under $\phi(X_t)$?

Simulation of diffusions

Continuous, strong Markov processes described by stochastic differential equation:

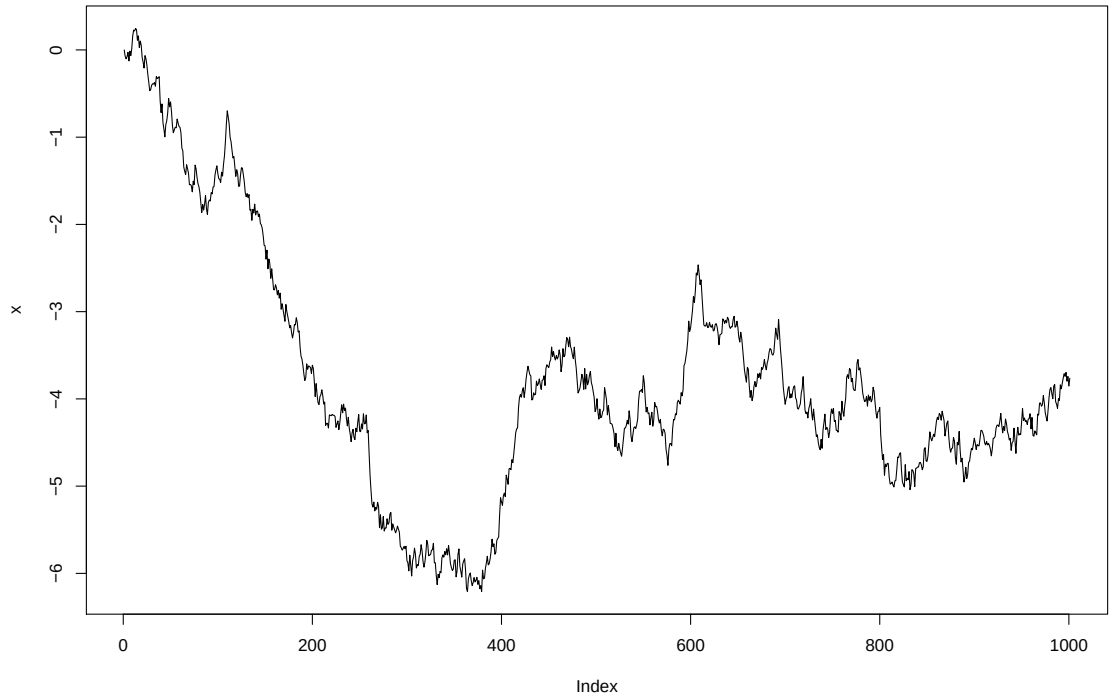
$$dX_t = \alpha(X_t)dt + \sigma(X_t)dB_t$$

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 $\mathbf{W}_0$ 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:

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

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 $A(x) = \int_0^x \alpha(u)du$ and 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

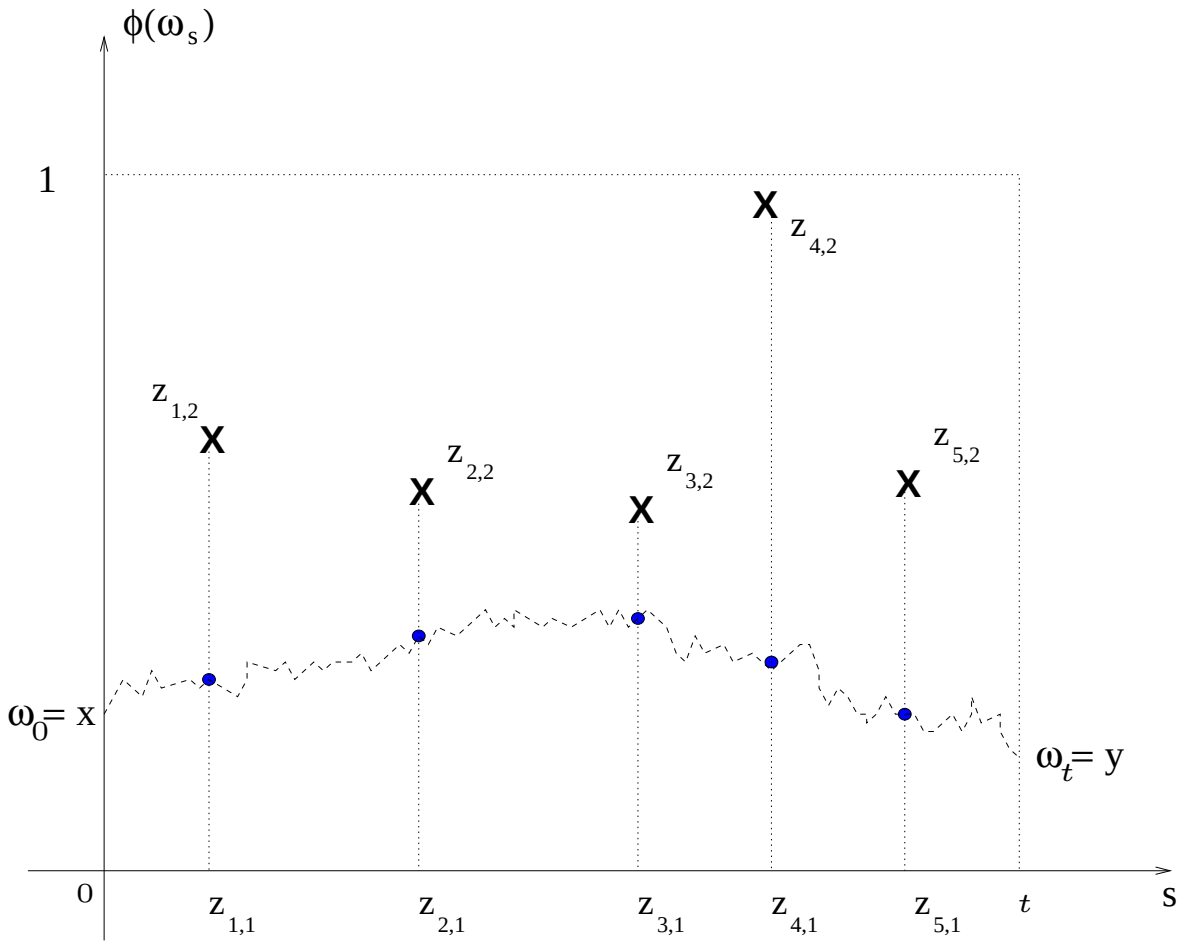
$$\frac{d\mathbf{P}}{d\mathbf{P}_0} \propto \exp\left\{-r \int_{s=0}^{1} \phi(X_s) ds\right\}.$$

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

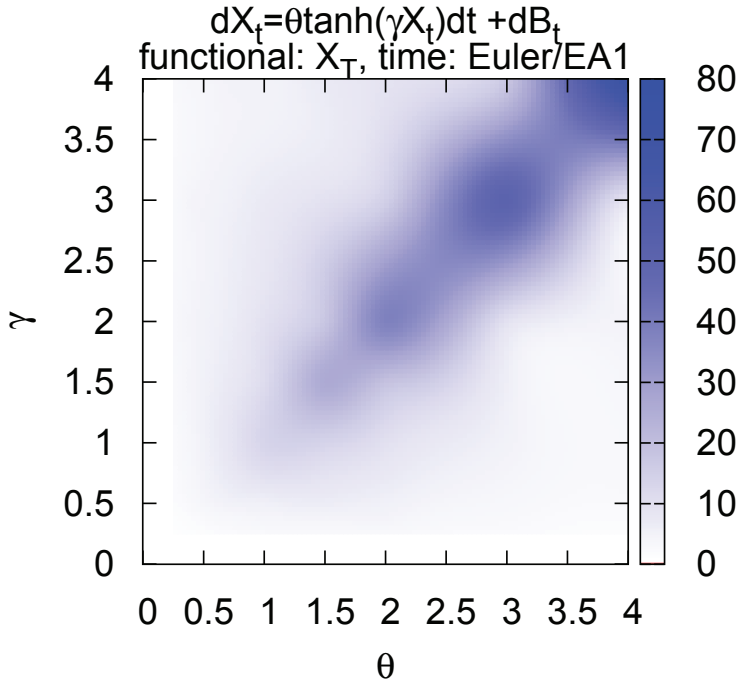
$$\mathbf{P}\left(\Phi \text{ is the empty configeration} = \exp\left\{-r \int_{0}^{1} \phi(X_s) ds\right\}\right).$$

The basic 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 $\{(0, B_0), (1, B_1), (U_i, B_{U_i}),\ 1 \leq i \leq n\}$.


$\phi(\omega_s)$
1
$z_{4,2}$
$z_{1,2}$
$z_{2,2}$
$z_{3,2}$
$z_{5,2}$
$\omega_0 = x$
$\omega_t = y$
0
$z_{1,1}$
$z_{2,1}$
$z_{3,1}$
$z_{4,1}$
$z_{5,1}$
t
s

Part of a simulation study



Some concluding remarks

More exotic algorithms needed to relax assumptions made.

Exact simulation of stochastic processes can be used very generally in statistical inference for stochastic processes. Current application areas include financial econometrics and infections disease modelling.

There are classes of stochastic processes for which the methodology works, but others where is intrinsically inapplicable. In such cases more complicated importance sampling methods can be adopted.

Retrospective simulation methods has very broad applications also in stochastic simulation. Many applications in Bayesian Statistics.

There is often no price for exactness! In fact exact methods frequently are more efficient than approximate alternatives.

Currently still a very active area of research.