i-like.org.uk

Intractable Likelihood

i-like Launch Day

January 31st, 2013

Modern likelihood

Cornerstone of modern statistical analysis:

- Maximum (pseudo/composite/quasi/penalised) likelihood
- Bayesian inference
- Hypothesis testing and model choice

The computational statistics revolution of the last 20-30 years has had many successes in making hard problems accessible to formal likelihood-based inference of some form, eg the EM algorithm, MCMC, Sequential Monte Carlo.

This has led to spectacular successes of likelihood based statistical analysis in just about every area in which data plays a role.

However the hardest of today's inference problems lie beyond the scope of current state-of-the-art likelihood methods.

Likelihood challenges

- Complex models:
  - designed to answer increasingly more intricate and subtle questions,
  - characterised by highly-dimensional (possibly infinite-dimensional) and highly structured parameter spaces.
- Massive data
  - Often likelihoods are too expensive to compute.
  - Both a problem and an opportunity!
- Optimisation of likelihood-based methods
  - Need to find ways assessing the statistical and computational properties of methods, and to optimise
  - incorporates online or adaptive methods

i-like.org.uk

A 5 year project funded through a *Programme Grant* from EPSRC to address these challenges.

Project will investigate wide-ranging applications. We will use three broad application areas where particularly challenging questions arise for likelihood:

- modelling of large scale networks (for example in ecology, commerce and bibliometrics),
- population genomics and genetics, and
- infectious disease epidemiology.

However applications will not be restricted to these areas.

i-like Postdoc positions available

5 postdoc positions in Intractable likelihood available starting in 2013 based at Warwick, Bristol, Lancaster, Oxford. Collaborative project involving Paul Fearnhead, David Firth, Christophe Andrieu, Chris Holmes, and me, Gareth Roberts.

Positions are for 2 years in the first instance with possible extension to 4 years.

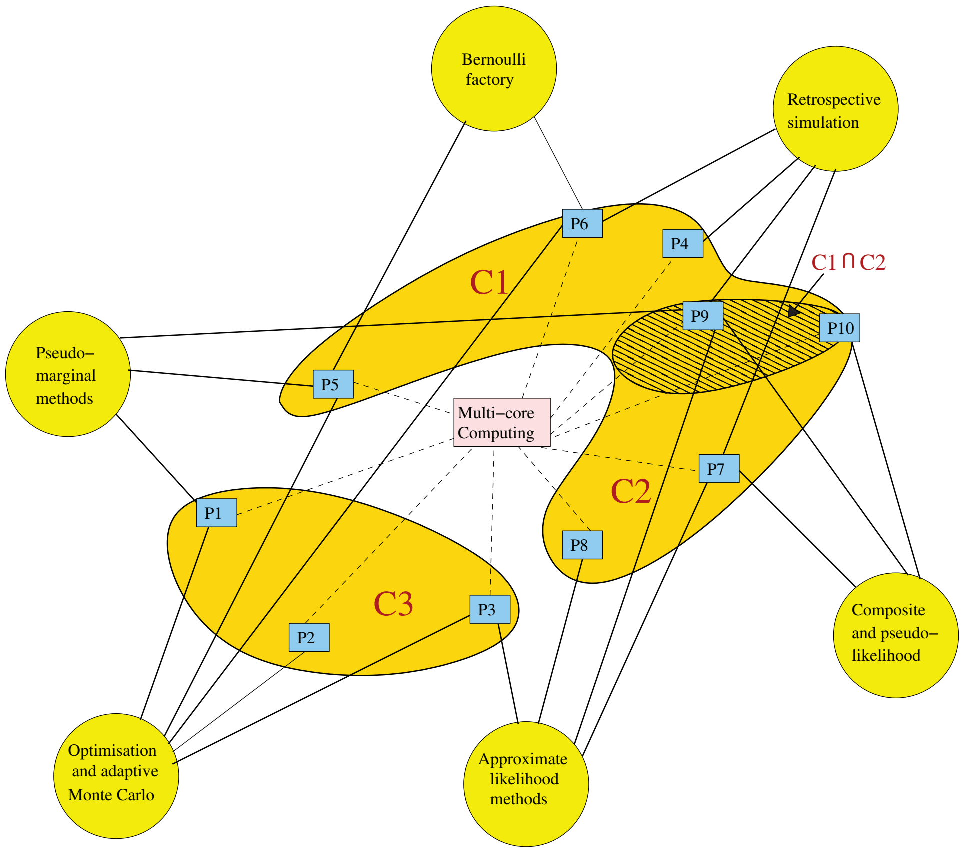
Christophe will say more later about these positions and how you can apply if you are interested.

Also we will be holding annual open workshops on Intractable Likelihood starting in 2014.

Project is really broad, and one aim is to stimulate further research in these areas.

What will *i-like* do?

Project is for 5 years. Fairly detailed plan for the first half of the project:



Further information can be found at `i-like.org.uk`

Retrospective Simulation

Gareth Roberts
University of Warwick

i-like Launch Day, Oxford, January 2013

This talk

Our ability to carry out likelihood-based inference is inevitably linked to our ability to carry out certain simulation tasks.

Problems within *i-like* often involve extremely high-dimensional (and sometimes infinite) parameter spaces, often entire trajectories of unobserved stochastic processes.

1. What is retrospective simulation?

2. Simulation of stochastic processes, eg diffusions

What is retrospective simulation?

It 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.

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".

Restrospective simulation has natural allies in the simulation game, for example catalytic perfect simulation and non-centering

Ex 1: The birth of retrospective simulation?

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

Who is the current heir to the throne?

1. Prince William
2. Prince Charles
3. David Beckham

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:

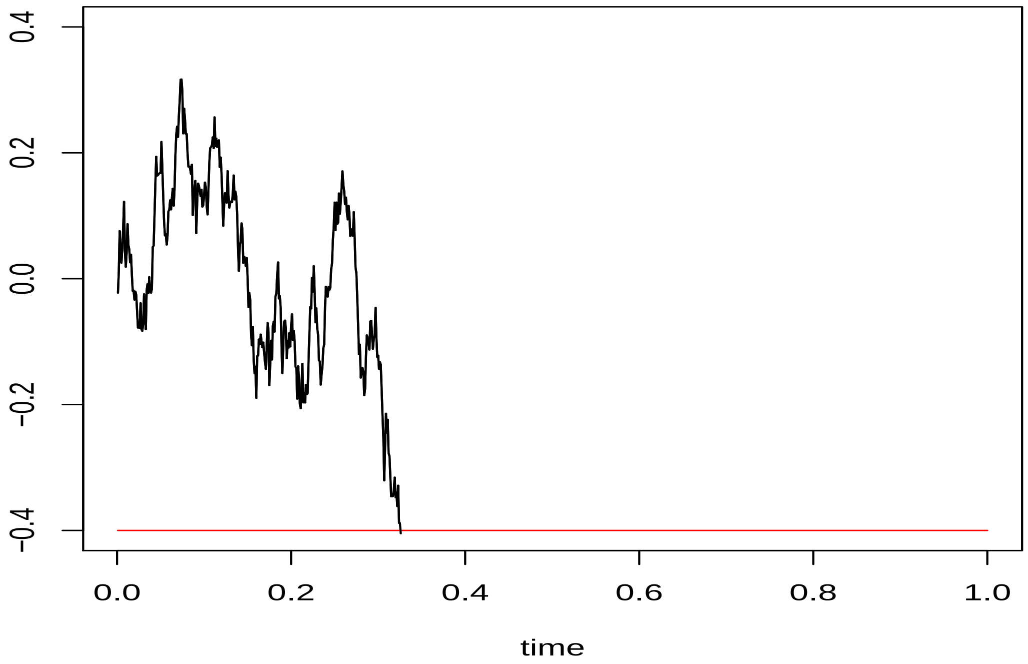
$$p_i^+ = \sum_{j=0}^{2i} a_j (-1)^j;$$

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

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

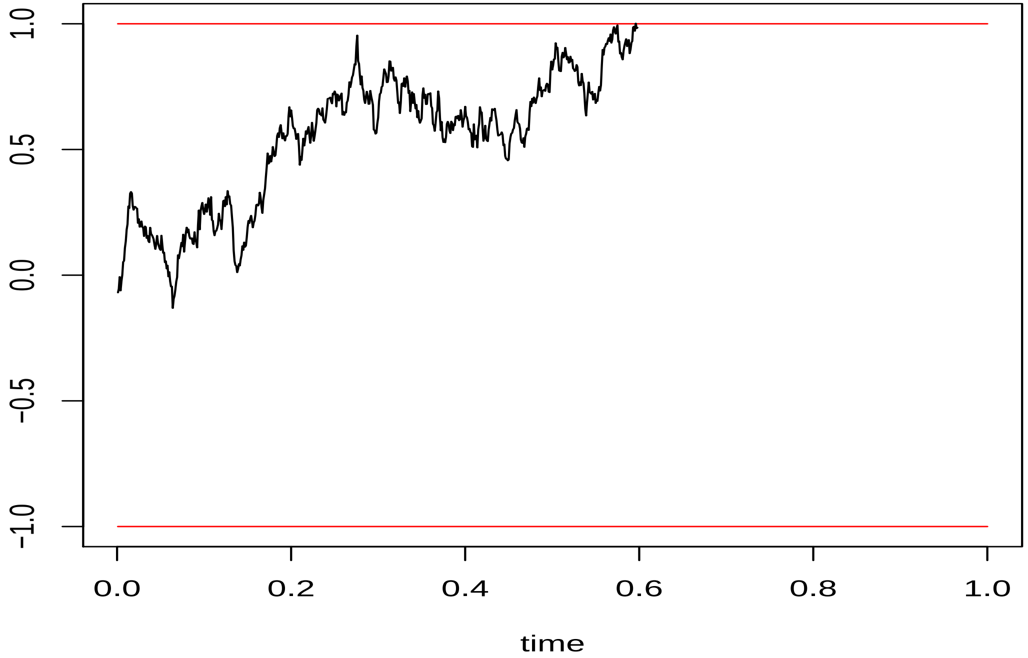
Let B_t be standard Brownian motion. Let $\tau_a = \inf\{t;\ B_t = a\}$



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

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

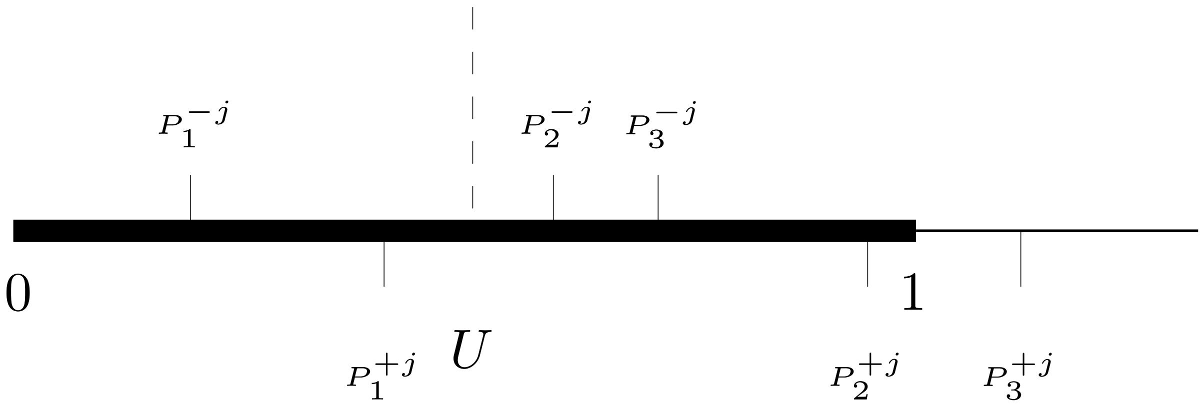
$$P_i^{+j} = \sum_{k=1}^{j} \frac{p_k}{s_j^-}$$

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

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

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

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 probabiltiy measure $\mathbf{P}_0$.

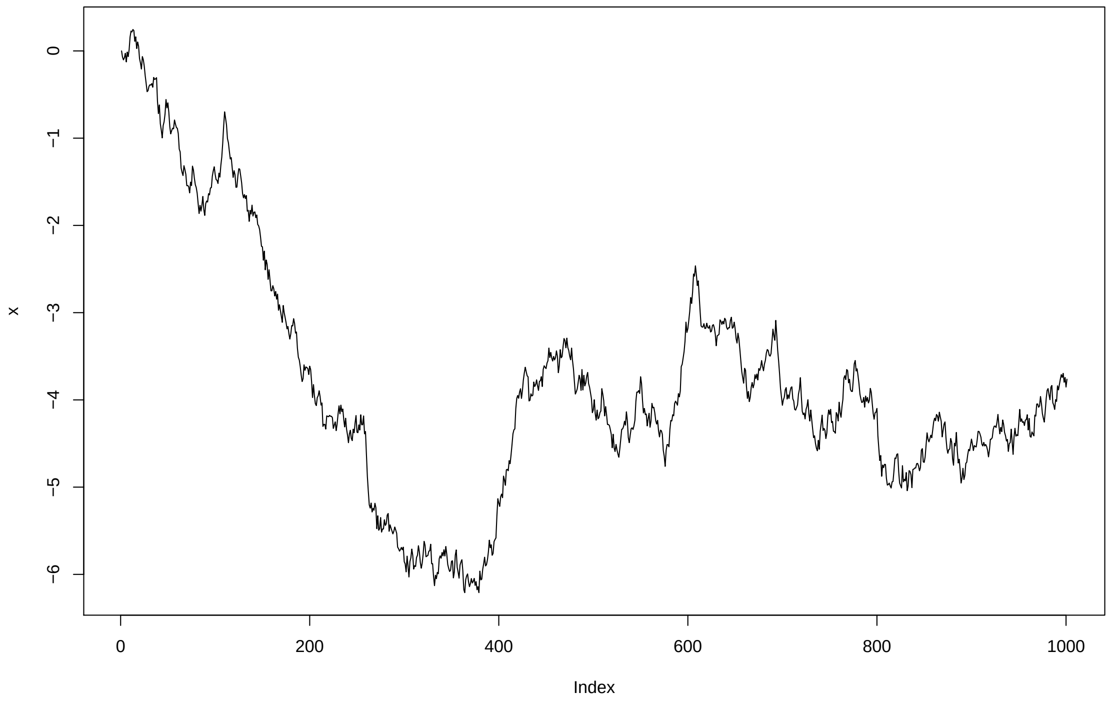
Suppose we are able to simulate from $\mathbf{P}_0$.

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-paramettrics.



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

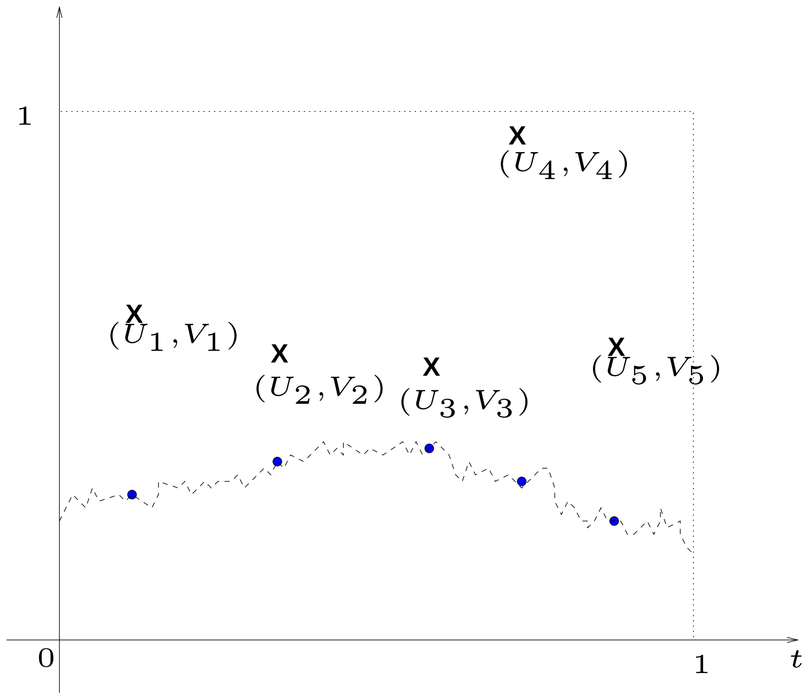
- Just storing all of X could require infinite storage capacity.
- Calculating $\int_0^1 \phi(X_s)ds$ is likely to require infinite computation

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 processof rate r on the unit square has no points on the epi graph $\{(u,v) \in [0,1]^2;\ v \le \phi(u)\}$.



Simulation of diffusions

Continuous, strong Markov processes described by stochastic differential equation:

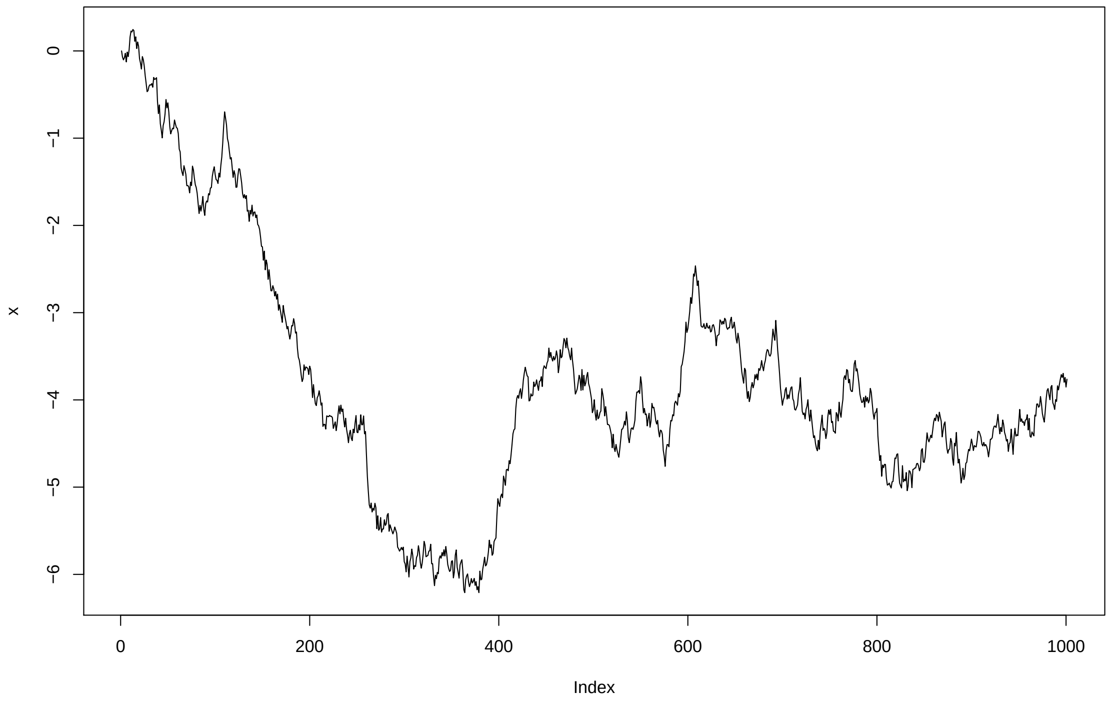
$$dX_t = \alpha(X_t)dt + \sigma(X_t)dB_t$$

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:

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

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 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 $\{0 \leq y \leq \phi(X_s),\ 0 \leq s \leq 1\}$. 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)

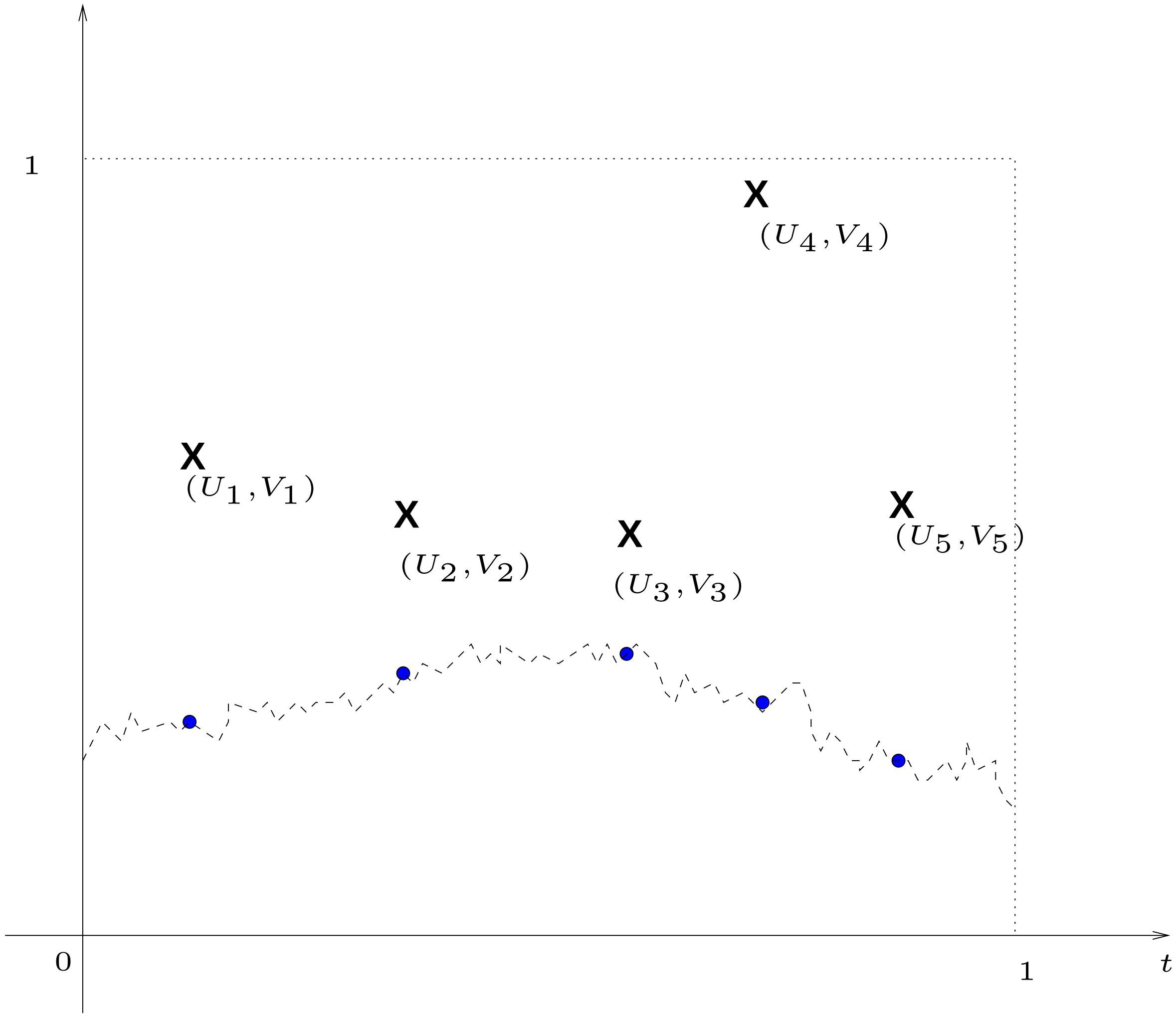
1. Set $B_0 = 0$. Simulate B_1 from (*)

2. Generate Poisson process of rate r on $[0,1] \times [0,1]$: $\Phi = \{(U_1, V_1), \ldots (U_n, V_n)\}$

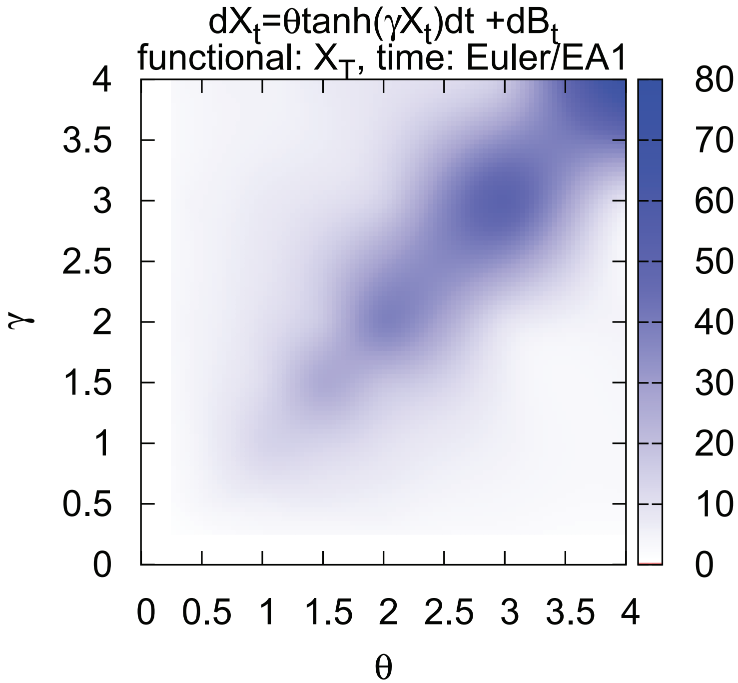
3. For each U_i, draw B_{U_i} from its appropriate Brownian bridge probabilities.

4. If $\phi(B_{U_i}) > V_i$ for ANY i, erase skeleton and go to (1).

5. Output the currently stored skeleton $\{(0, B_0), (1, B_1), (U_i, B_{U_i}),\ 1 \leq i \leq n\}$.


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

Part of a simulation study



Summary

- Retrospective simulation is in principle simple, though the devil is in the detail!
- Using retrospective simulation to do exact simulation is not necessarily computationally costly in comparison to approximation methods.
- Many examples of retrospective simulation being used in statistics, eg inference for diffusions and Bayesian non-parametrics.
- So retrospective simulation methods can be expected to feature strongly in the *i-like* programme as it develops.

Some relevant references

1. Beskos A. Papaspiliopoulos O. and Roberts G.O. Retrospective Exact Simulation of Diffusion Sample Paths with Applications, *Bernoulli*, 12, 6, 1077-1098, 2006.

2. Beskos A. Papaspiliopoulos O. Roberts G.O. and Fearnhead P. Retrospective Exact Simulation of Diffusion Sample Paths with Applications, Exact and computationally efficient likelihood–based estimation for discretely observed diffusion processes, *J. Roy Stat Soc.*, B, 68, 3, (with discussion), 333-382, 2006.

3. Casella B. and Roberts G.O. Exact Monte Carlo simulation of killed diffusions, *Adv. Appl. Prob.*, **40**, 273–291, 2008.

4. Beskos A. Papaspiliopoulos O. and Roberts G.O. A factorisation of diffusion measure and finite sample path constructions, *Methodology and Computing in Applied Probability*, **10**, 1, 85–104, 2008.

5. Papaspiliopoulos O. and Roberts G.O. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models,

Biometrika, **95**, 169–186, 2008.

6. Fearnhead P. Papaspiliopoulos O. and Roberts G.O. Particle filters for partially-observed diffusions, *Journal of the Royal Statistical Society*, Series B, **70**, 755-777, 2008.

7. Beskos A. Papaspiliopoulos O. and Roberts G.O. Monte Carlo Maximum Likelihood for discretely observed diffusions, *Annals of Statistics*, **37**, 1, 223-245, 2009.

8. Fearnhead P. Papaspiliopoulos O. Roberts G.O. and Stuart A. Wald random weighted importance sampling with application to continuous time filtering, *J. Roy. Statist. Soc., series B*, **72**, 4, 497–512, 2010.

9. Papaspiliopoulos O. Yau C. Holmes, C. and Roberts G.O. Bayesian nonparametric Hidden Markov Models with applications in genomics, *J. Roy. Statist. Soc., series B*, **73**, 1, 35–57, 2011.

10. Casella B. and Roberts G.O. Exact simulation of jump-diffusion

processes with Monte Carlo applications, *Methodology and Computing in Applied Probability*, **13**, 3, 449–473, 2011.

11. Peluchetti, S.and Roberts G.O. A study of the efficiency of exact methods of diffusion simulation, in *Monte Carlo and Quasi-Monte Carlo Methods 2010*, 161–187, 2012.

12. Beskos, A. Peluchetti S.and Roberts G.O. ϵ strong simulation of the Brownian path, *Bernoulli*, **18** , 4, 1223–1248, 2012.

13. Papaspiliopoulos, O.Roberts, G.O. and Stramer, O. Importance sampling techniques for estimation of diffusion models, in *Statistical Methods for Stochastic Differential Equations*, Chapman and Hall, 124, 311-340, 2012.

14. Sermaides, G. Papaspilipoulos, O. Roberts, G.O. Beskos, A.and Fearnhead, P. Markov chain Monte Carlo for exact inference for diffusions, to appear in *Scandinavian Journal of Statistics*, 2013.