Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling

Hermann Singer

Diskussionsbeitrag Nr. 497

November 2016

Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung
FernUniversität in Hagen
Universitätsstraße 41
58084 Hagen
http://www.fernuni-hagen.de/ls_statistik/
Sekretariat.Statistik@FernUni-Hagen.de

Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling

Hermann Singer
FernUniversität in Hagen *

November 11, 2016

Abstract

Continuous time models are well known in sociology through the pioneering work of Simon (1952); Coleman (1968); Doreian and Hummon (1976, 1979) and others. Although they have the theoretical merit in modeling time as a flowing phenomenon, the empirical application is more difficult in comparison to time series models. This is in part due to the difficulty in computing likelihood functions for sampled, discrete time measurements (daily, weekly etc.), as they occur in empirical research.

With large sampling intervals, one cannot simply replace differentials by differences, since then one obtains strongly biased estimates of structural parameters. Instead one has to consider the exact transition probabilities between the times of measurement. Even in the linear case, these probabilities are nonlinear functions of the structural parameter matrices with respective identification and embedding problems (Hamerle et al.; 1991).

For nonlinear systems, additional problems occur due to the impossibility of computing analytical transition probabilities for most models. There are competing numerical methods based on nonlinear filtering, partial differential equations, integral representations, Monte Carlo and Bayesian approaches.

We compute the likelihood function of a nonlinear continuous-discrete state space model (continuous time dynamics of latent variables, discrete time

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany, hermann.singer@fernuni-hagen.de

Paper presented at the 9th International Conference on Social Science Methodology (RC33), 11.-16. September 2016, Leicester, UK. Session 19: Estimation of Stochastic Differential Equations with Time Series, Panel and Spatial Data.

noisy measurements) by using a functional integral representation. The unobserved state vectors are integrated out in order to obtain the marginal distribution of the measurements.

The infinite-dimensional integration is evaluated by Monte Carlo simulation with importance sampling. Using a Langevin equation with Metropolis mechanism, it is possible to draw random vectors from the exact importance distribution, although the normalization constant (the desired likelihood function) is unknown. We discuss several methods of estimating the importance distribution. Most importantly, we obtain smooth likelihood surfaces which facilitates the usage of quasi Newton algorithms for determinating the ML estimates. The proposed Monte Carlo method is compared with Kalman filtering and analytical approaches using the Fokker-Planck equation.

More generally, one can compute functionals of diffusion processes such as option prices or the Feynman-Kac formula in quantum theory.

Key Words:

- Stochastic Differential Equations - Nonlinear continuous-discrete state space model - Simulated Maximum Likelihood - Langevin Importance Sampling

1 Introduction

Theoretical work in sociology and economics frequently uses dynamical specifications in continuous time, formulated in the language of deterministic or stochastic differential equations. On the other hand, econometric estimation methods for the structural parameters of these equations are often formulated in discrete time (time series and panel models). This is mostly due to the fact, that measurements are usually given at discrete time points (daily, weekly, monthly etc.). As long as the sampling intervals are small, there seems to be no problem, since differential equations can be discretized (Kloeden and Platen; 1999). However, for large intervals, these approximations involve large errors. Therefore, one should distinguish between a dynamically relevant (discretization) interval δt and a measurement interval Δt. Conventional time series and panel analysis can be viewed as setting these intervals equal, whereas differential equation models consider the limit $\delta t \to 0$.

In this paper we follow the intermediate approach, that δt is so small, that the involved approximations are reasonably good, but the computational demand is tractable. The states between the measurements are treated as missing. Therefore a measurement model is introduced, which also can accomodate errors of measurement (errors in variables) and unobserved components. This concept can be used both in a Kalman filter or a structural equations context (Singer; 1995, 2007, 2012).

In the case of nonlinear dynamics, one can use recursive filter equations, which allow both the computation of estimates of latent (nonobserved) states and the likelihood function, thus permitting maximum likelihood estimation. Unfortunately, in Monte Carlo implementations of this approach, the simulated likelihood is not always a smooth function of the parameters, thus Newton-type optimization algorithms may run into difficulties (cf. Pitt; 2002; Singer; 2003).

In this paper, we use alternatively a nonrecursive approach to compute the likelihood function. The probability density of the measurements is obtained after integrating out all latent, unobserved states. This integration is achieved by a Markov chain Monte Carlo (MCMC) method, called Langevin sampling (Roberts and Stramer; 2002). The efficiency of the integration is improved by using importance sampling (cf. Durham and Gallant; 2002). Although the importance density is only known up to a factor, one can draw a random sample from this distribution and obtain an estimate therof. This will lead to a variance reduced MC computation of the desired likelihood function. An analogous approach can be used to compute functionals occuring in finance and quantum mechanics.

2 State Space Models

2.1 Nonlinear continuous/discrete state space model

The nonlinear continuous/discrete state space model (Jazwinski; 1970) consists of a dynamic equation and a measurement equation

$$\begin{aligned} dY(t) &= f(Y(t),t)dt + g(Y(t),t)dW(t) && (1) \\ Z_i &= h(Y_i,t_i) + \epsilon_i && (2) \\ i &= 0,...,T, \end{aligned}$$

where the states $Y(t) \in \mathbb{R}^p$ can be measured only a certain discrete time points t_i. This is the usual case in empirical research. In the equations, we encounter

- nonlinear drift and diffusion functions f, g and an output function h which depend on a parameter vector $\psi \in \mathbb{R}^u$, i.e. $f = f(Y(t), t, \psi)$
- and use Itô calculus in the case of nonlinear diffusion functions $g(Y)$.
- Spatial models for the random field $Y(x,t)$ can be treated by setting $Y_n(t) = Y(x_n,t), x_n \in \mathbb{R}^d, n = 1,...,p$ (see fig. 1). Note that not all $Y(x_n,t)$ are necessarily observable (see Singer; 2011).

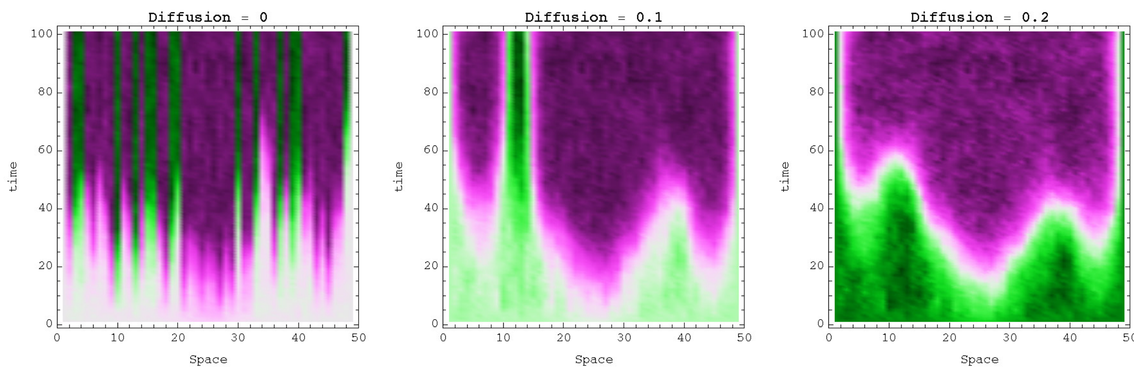

Figure 1: Nonlinear spatial model (Ginzburg-Landau equation) for $Y(x_n, t)$.

2.2 Linear stochastic differential equations (LSDE)

An important special case is the system of linear stochastic differential equations (LSDE, Singer; 1990) with initial condition $Y(t_0)$ and solution

$$\begin{aligned} dY(t) &= AY(t)dt + GdW(t) \\ Y(t) &= e^{A(t-t_0)}Y(t_0) + \int_{t_0}^{t} e^{A(t-s)}GdW(s). \end{aligned}$$

In the equations, we use

- the Wiener process $W(t, \omega) \in \mathbb{R}^r$, a continuous time random walk (Brownian motion),
- from which we can derive the Gaussian white noise $dW/dt = \zeta(t)$ with autocovariance $E[\zeta(t)\zeta'(s)] = \delta(t-s)I_r$.
- $A \in \mathbb{R}^{p,p}$ and $G \in \mathbb{R}^{p,r}$ are called drift matrix and diffusion matrix, respectively.
- The exact discrete model (EDM) valid at times $t = t_{i+1}, t_0 = t_i$ is due to Bergstrom (1976, 1988)

2.3 Exact discrete model (EDM)

At the times of measurement ($t = t_{i+1}, t_0 = t_i$), we obtain a restricted VAR(1) autoregression (Bergstrom; 1976, 1988)

$$Y_{i+1} = e^{A(t_{i+1}-t_i)}Y_i + \int_{t_i}^{t_{i+1}} e^{A(t_{i+1}-s)}GdW(s), \tag{3}$$

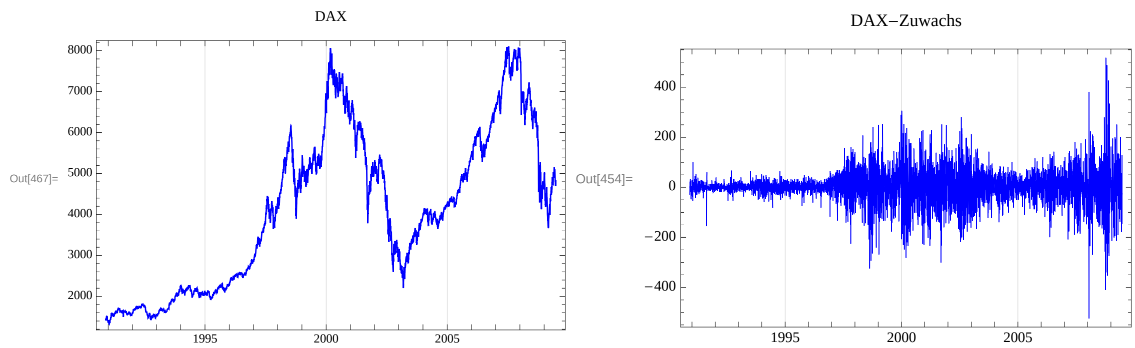

Figure 2: German stock index (DAX)

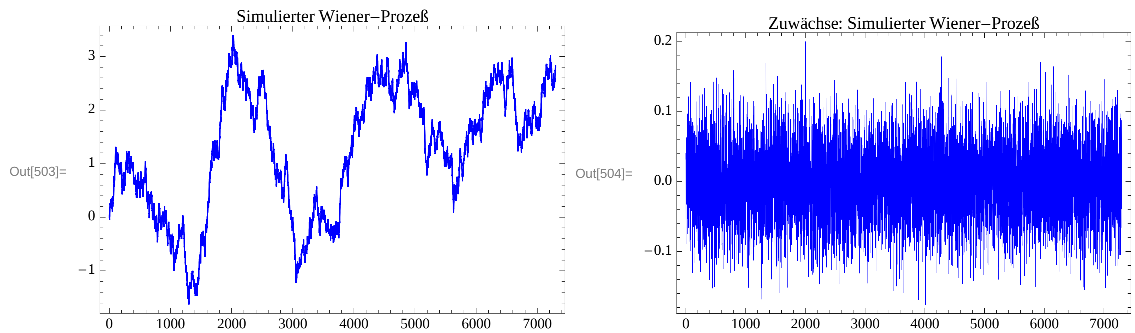

Figure 3: Top: Simulated Wiener process (random walk). Bottom: increments.

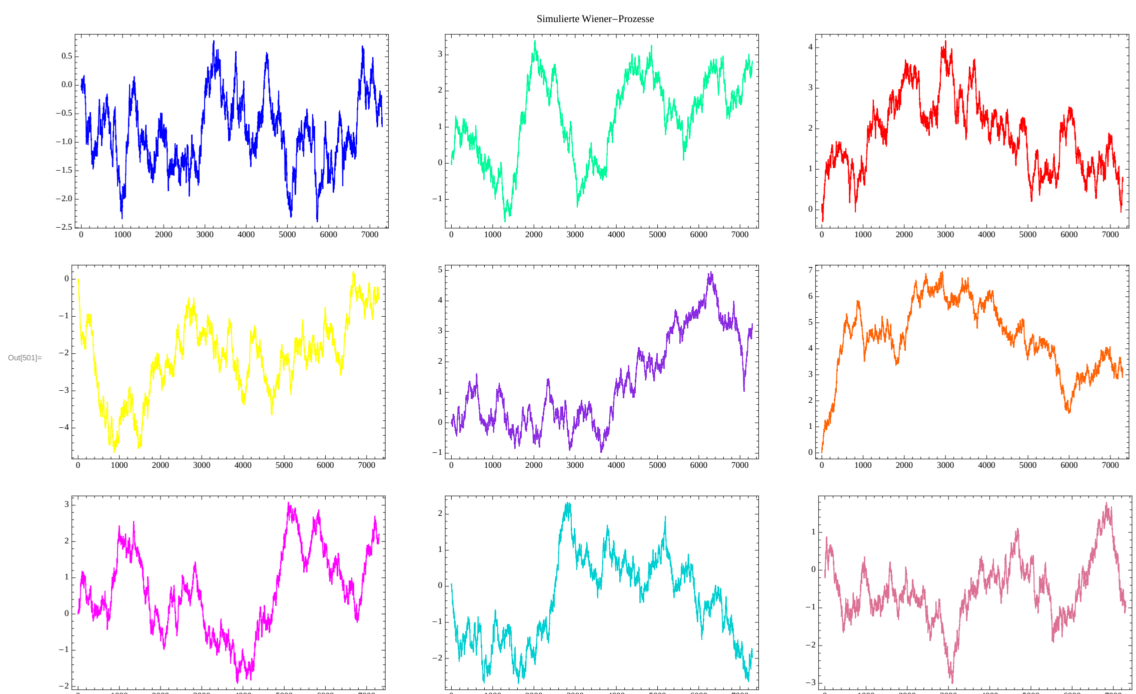

Figure 4: Simulated Wiener processes

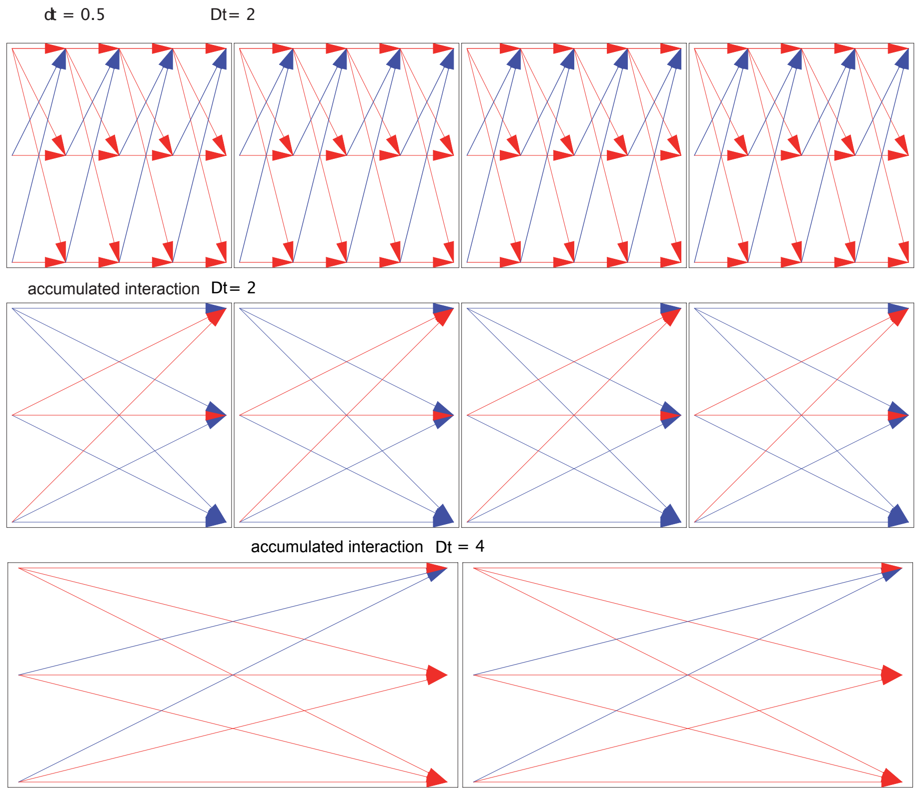

Figure 5: 3-variables-model: Product representation of matrix exponential within the measurement interval $\Delta t = 2$. Latent states η_j, discretization interval $\delta t = 2/4 = 0.5$ (Singer; 2012)

abbeviated as

$$Y_{i+1} = \Phi(t_{i+1}, t_i) Y_i + u_i$$

In this equation, we use the notation

- Φ (fundamental matrix of the system),
- $Y_i := Y(t_i)$ are the 'sampled' measurements and
- $\Delta t_i := t_{i+1} - t_i$ is the sampling (measurement) interval.

3 State and parameter estimation

3.1 Exact continuous-discrete filter

Central to the treatment of dynamic state space models is the recursive estimation of the latent states $Y(t)$. We describe their probability distribution by conditional

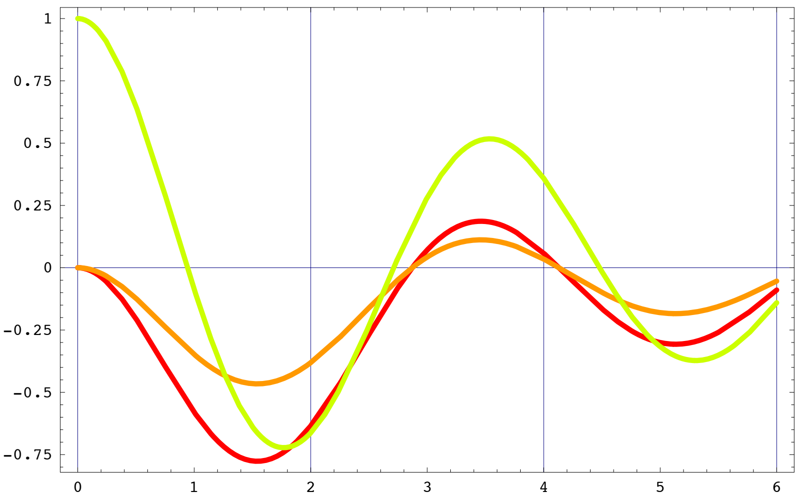

Figure 6: 3-variables-model: time dependency of the discrete time parameter matrices $A^*(\Delta t) = \exp(A\Delta t)$ form the sampling interval Δt. Matrix elements A^*_{12}, A^*_{21}, A^*_{33} (red, yellow, green).

densities $p(y_i|Z^i)$, given measurements up to time t_i ($i = 0, \ldots, T-1$). Then one gets the recursive sequence (Bayes filter)

Time update (prediction):

$$p(y_{i+1}|Z^i) = \int p(y_{i+1}|y_i, Z^i)p(y_i|Z^i)dy_i$$

Measurement update (Bayes formula):

$$p(y_{i+1}|Z^{i+1}) = \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)}$$

Conditional Likelihood:

$$p(z_{i+1}|Z^i) = \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)dy_{i+1}$$

with the nomenclature

- $p(y_{i+1}|Z^i)$: a priori probability density
- $p(y_{i+1}|Z^{i+1})$: a posteriori probability density; including a new measurement Z_{i+1}
- $Z^i = \{Z(t_j)|t_j \leq t_i\}$: observations up to time t_i, $Z_i := Z(t_i)$
- $p(z_{i+1}|Z^i)$: (conditional) likelihood function of observation $Z_{i+1} = z_{i+1}$.

From this, one can compute the so called prediction error decomposition (recursive likelihood function of all observations; Schweppe 1965)

$$p(z_T, ..., z_0; \psi) = \prod_{i=0}^{T} p(z_{i+1}|Z^i)\; p(z_0). \tag{4}$$

4 Parameter Estimation

In empirical applications in the social sciences, usually the parameter vector ψ is unknown and cannot be measured separately from the observations Z_i. One can either use (4) for maximum likelihood (ML) parameter estimation (Singer; 1995, 2015) or utilize a nonrecursive formula based on the representation

$$p(z_T, ..., z_0; \psi) \;=\; \int p(z_T, ..., z_0|y_T, ..., y_0) p(y_T, ..., y_0) dy_0 ... dy_T. \tag{5}$$

Using this formula, we have

- a high dimensional integration over latent variables
- and a smooth dependence on the parameter vector ψ: $L(\psi) = p(z; \psi)$.
- However, the joint density $p(y_T, ..., y_0) = p(y_T|y_{T-1})....p(y_1|y_0)p(y_0)$ is in most cases not explicitly known, due to the (long) sampling interval Δt_i.
- In particular, the transition density $p(y_{i+1}|y_i) := p(y_{i+1}, \Delta t_i|y_i)$ is difficult to compute.

One could use the method of Aït-Sahalia (2002, 2008)[1] or Li (2013) to obtain an asymptotic expansion of p. A more simple approach is the so called Euler transition density, which is valid over short time intervals, or the so called local linearization (LL) method, which is exact for linear systems (Shoji and Ozaki; 1998b). The latter methods have the advantage, that the density approximation integrates to unity (cf. Stramer et al.; 2010).

In this paper, we compute the likelihood function as input to a quasi-Newton optimization algorithm, whereas Stramer et al. (2010) directly sample from the posterior density. The latter algorithm may run into difficulties for small δt, since the quadratic variation $dy^2 = g(y, \psi)^2 dt$ of the latent states contains the diffusion parameters. A remedy is the method of transformations (Roberts and Stramer; 2001; Dargatz; 2010) or the usage of good approximations of $p(y_{i+1}, \Delta t_i|y_i)$ over the finite sampling interval Δt_i.

[1] only for reducible diffusions

In the likelihood approach (5), the sampling problem for the diffusion parameters does not occur. We perform the integration by using additional latent variables η_j (see fig. 5)

$$\begin{aligned} y_T &= \eta_J, ..., \eta_0 = y_0 \\ y_i &= \eta_{j_i} \\ j &= 0, ..., J = (t_T - t_0)/\delta t,\ i = 0, ..., T, \end{aligned}$$

which is a data augmentation algorithm (Tanner and Wong; 1987; Tanner; 1996).

4.1 Integration

Inserting the latent states η_j into equation (5) we obtain the integral representation

$$\begin{aligned} p(z_T, ..., z_0) &= \int p(z_T, ..., z_0|\eta_J, ..., \eta_0)p(\eta_J, ..., \eta_0)d\eta && (6) \\ &= E\Big[p(z_T, ..., z_0|\eta_J, ..., \eta_0)\Big] && (7) \end{aligned}$$

with notation $\eta_{j_i} = y(t_i) := y_i$ at the measurement times t_i. Now we have an even higher dimensional integration over latent variables, which is performed by simulation using Markov Chain Monte Carlo (MCMC).

For small discretization interval δt one can use the so called Euler density

$$p(\eta_{j+1}, \delta t|\eta_{j+1}) \approx \phi(\eta_{j+1}; \eta_j + f_j\delta t, \Omega_j\delta t),$$

setting $f_j := f(\eta_j, \tau_j), \Omega_j := (gg')(\eta_j, \tau_j)$. δt is typically much smaller than the measurement interval Δt (Singer; 1995). At this point, if better approximations for p (e.g. Li; 2013, loc. cit.) are used, we can choose a larger δt leading to a smaller dimension of the latent state.

Replacing the expectation value (7) by a mean value, we obtain a MC estimator for the desired likelihood function, i.e.

$$\hat{p}(z_T, ..., z_0) = L^{-1}\sum_l p(z_T, ..., z_0|\eta_{Jl}, ..., \eta_{0l}). \quad (8)$$

However, this estimator is extremely inefficient, since most samples (trajectories) $(\eta_{Jl}, ..., \eta_{0l})$ yield very small contributions $p(z_T, ..., z_0|\eta_{Jl}, ..., \eta_{0l})$. One may imagine, that most trajectories are far from the given measurements.

4.1.1 Importance Sampling

We use the well known method of importance sampling (Kloeden and Platen; 1999) to get a variance reduced MC integration of the form

$$p(z_T, ..., z_0) = \int p(z|\eta)\frac{p(\eta)}{p_2(\eta)}p_2(\eta)d\eta \quad (9)$$

where

- p_2 is the so called importance density with optimal choice:

- $p_{2,optimal} = \frac{p(z|\eta)p(\eta)}{p(z)} = p(\eta|z)$.

- the integration (9) is performed by averaging over a random field with equilibrium distribution $p_{2,optimal}$, i.e.

  $\eta = \eta(t, u, \omega)$: $t =$ true time, $u =$ simulation time.

  Actually, we use a finite dimensional approximation $\eta(u) = \eta_j(u) = \eta(\tau_j, u)$, $j = 0, ..., J$, on a time grid $\tau_j = j\delta t$.

However, $p(z)$ is unknown (it is the desired quantity).

4.2 Langevin Sampling

Fortunately, we can sample from $p_{2,optimal}$, although $p(z)$ is unknown.

Using the so called Langevin equation (Langevin; 1908; Roberts and Stramer; 2002)

$$d\eta(u) = (\partial_\eta \log p(\eta|z))(\eta(u))du + \sqrt{2}\, dW(u), \tag{10}$$

we can simulate states $\eta(u)$ which stem from the desired distribution $p_{2,optimal}$.

If the dynamical system described by (10) is in equilibrium (stationary state) we get the results:

- The stationary distribution of conditional latent states $\eta(u)$ is given by

$$p_{stat}(\eta) = p(\eta|z) = \lim_{u\to\infty} p(\eta, u).$$

- It can be written as $p_{stat}(\eta) = e^{-\Phi(\eta)} = p(\eta|z)$

- The drift function in (10) is the negative gradient of a 'potential'

$$\Phi(\eta) \;:=\; -\log p(\eta|z) \tag{11}$$

$$-\partial_\eta \Phi(\eta) \;=\; \partial_\eta[\log p(z|\eta) + \log p(\eta) - \log p(z)]. \tag{12}$$

  Thus we can sample from $p(\eta|z)$ and $p(z)$ is not needed.

- As a by-product, optimal nonlinear smoothing

  $\eta(u) \sim p(\eta|z)$ in equilibrium $u \to \infty$, can be performed.

The concept of a 'potential' (e.g. electrostatic or gravitational) is borrowed from physics and we can understand equation (10) as the overdamped random movement of a fictious high dimensional object in a force field given by the negative gradient (Nelson; 1967). Of course, the coordinate vector $\eta = (\eta_0, ..., \eta_J); J = (t_T - t_0)/\delta t$ is infinite dimensional in the continuuum limit $\delta t \to 0$. For an analytical computation of the drift function in (10), see Reznikoff and Vanden-Eijnden (2005); Hairer et al. (2007); Singer (2016).

4.3 Simulated Likelihood

Summarizing, we compute the simulated likelihood using variance reduced MC-integration by the formula

$$\hat{p}(z_T, ..., z_0) \quad = \quad L^{-1} \sum p(z|\eta_l) \frac{p(\eta_l)}{p_{2,optimal}(\eta_l)} \tag{13}$$

where

- $\eta_l \sim p(\eta|z)$, if the Langevin equation (10) is in an stationary (equilibrium) state.

- We draw optimal samples $\eta_l = \eta(u_l)$ from the Euler-discretized Langevin equation (including a Metropolis-Hastings mechanism). This ensures that approximation errors are compensated (Roberts and Stramer; 2002). Actually, we use an Ozaki scheme which is exact for linear systems (Ozaki; 1985).

- However, $p_{2,optimal} = \frac{p(z|\eta)p(\eta)}{p(z)} = p(\eta|z)$ cannot be computed, because $p(z)$ is unknown.

We attempt to estimate $p_2 = p(\eta|z)$ from the simulated data $\eta_l \sim p(\eta|z)$.

Estimation of the importance density can be performed in several ways, including:

- Use known (suboptimal) reference density

$$p_2 = p_0(\eta|z) = p_0(z|\eta)p_0(\eta)/p_0(z)$$

- Use kernel density estimate

$$\hat{p}(\eta|z) = L^{-1} \sum_l k(\eta - \eta_l; \text{smoothing parameter}) \tag{14}$$

  Problem:

  - high dimensional state η,
  - no structure imposed on $p(\eta|z)$

4.4 Estimation of importance density

We use the Markov structure of the (Euler discretized) state space model

$$\begin{aligned} \eta_{j+1} &= f(\eta_j)\delta t + g(\eta_j)\delta W_j \\ z_i &= h(y_i) + \epsilon_i \end{aligned}$$

and the Bayes formula

$$p(\eta|z) = p(\eta_J|\eta_{J-1}, ..., \eta_0, z)\ p(\eta_{J-1}, ..., \eta_0|z).$$

Now it can be shown that η_j is a conditional Markov process

$$p(\eta_{j+1}|\eta_j, ..., \eta_0, z) = p(\eta_{j+1}|\eta_j, z). \tag{15}$$

To see this, we use the conditional independence of the past $z^i = (z_0, ..., z_i)$ and future $\bar{z}^i = (z_{i+1}, ...z_T)$ given η^j. One obtains

$$\begin{aligned} p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i) &= p(\eta_{j+1}|\eta^j, \bar{z}^i) = p(\eta_{j+1}|\eta_j, \bar{z}^i) \\ p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i) &= p(\eta_{j+1}|\eta_j, \bar{z}^i) \\ j_i \leq j &< j_{i+1} \end{aligned}$$

since

(i) the transition density $p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i)$ is independent of past measurements, given the past true states, and only the last state η_j must be considered (Markov process).

(ii) the transition density $p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i)$ is independent of past measurements.

Thus we have $p(\eta_{j+1}|\eta^j, z^i, \bar{z}^i) = p(\eta_{j+1}|\eta_j, z^i, \bar{z}^i)$ ■

For the estimation of importance density (15), two methods are discussed:

4.4.1 Euler transition kernel with modified drift

- Euler density (discretization interval δt)

$$p(\eta_{j+1}, \delta t|\eta_j) \approx \phi(\eta_{j+1}; \eta_j + f_j\delta t, \Omega_j\delta t)$$

- modified drift and diffusion matrix (conditional Euler density)

$$p(\eta_{j+1}, \delta t|\eta_j, z) \approx \phi(\eta_{j+1}; \eta_j + (f_j + \delta f_j)\delta t, (\Omega_j + \delta\Omega_j)\delta t) \tag{16}$$

- nonlinear regression for δf_j and $\delta\Omega_j$: parametric and nonparametric

4.4.2 Kernel density estimation

conditional transition density

$$p(\eta_{j+1}, \delta t|\eta_j, z) = \frac{p(\eta_{j+1}, \eta_j|z)}{p(\eta_j|z)} \tag{17}$$

- estimate joint density $p(\eta_{j+1}, \eta_j|z)$ and $p(\eta_j|z)$ with kernel density estimates
- variant: use Gaussian $\phi(\eta_{j+1}, \eta_j|z)$ and $\phi(\eta_j|z)$ instead.

In both cases, data $\eta_{jl} = \eta(\tau_j, u_l) \sim p(\eta|z)$ are drawn from Langevin equation (10).

5 Examples

5.1 Geometrical Brownian motion (GBM)

The SDE

$$dy(t) = \mu y(t)dt + \sigma y(t)\, dW(t)$$

is the best known model for stock prices. It was used by Black and Scholes (1973) for modeling option prices, contains a multiplicative noise term $y\, dW$ and is thus bilinear. The form

$$dy(t)/y(t) = \mu dt + \sigma\, dW(t)/dt$$

shows, that the simple returns are given by a constant value μdt plus white noise.

In summary, we have the properties:

- log returns: set $x = \log y$, use Itô's lemma
  $dx = dy/y + 1/2(-y^{-2})dy^2 = (\mu - \sigma^2/2)dt + \sigma dW$
- exact solution: multiplicative exact discrete model
  $$y(t) = y(t_0)e^{(\mu-\sigma^2/2)(t-t_0)+\sigma[W(t)-W(t_0)]}$$
- exact transition density (log-normal distribution)
  $$p(y, t|y_0, t_0) = y^{-1}\phi\left(\log(y/y_0); (\mu - \sigma^2/2)(t - t_0), \sigma^2(t - t_0)\right)$$

The model was simulated using $\mu = 0.07, \sigma = 0.02$ and $\delta t = 1/365$. Only monthly data were used (fig. 7). We obtain a smooth likelihood surface with small approximation error (fig. 9). Clearly, the usage of the full kernel density (14) yields bad results (fig. 10). On the other hand, the representation (17) works very well (fig. 9). One can also use a gaussian density or a linear GLS estimation of the drift correction δf_j and diffusion correction $\delta\Omega_j$ (see eqn. 16). If the diffusion matrix is not corrected, biased estimates occur (fig. 13).

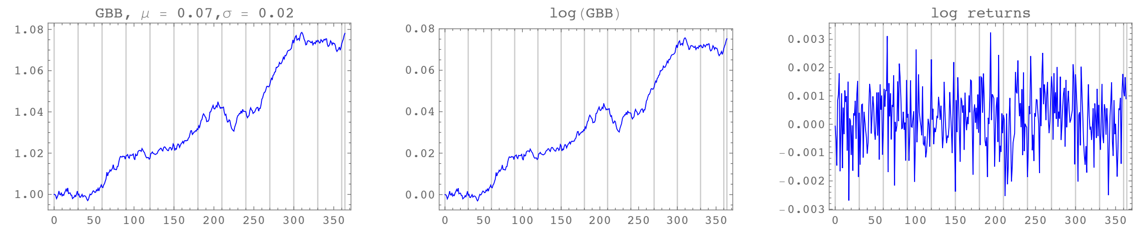

Figure 7: Geometrical Brownian motion: Trajectory and log returns. Vertical lines: measurement times.

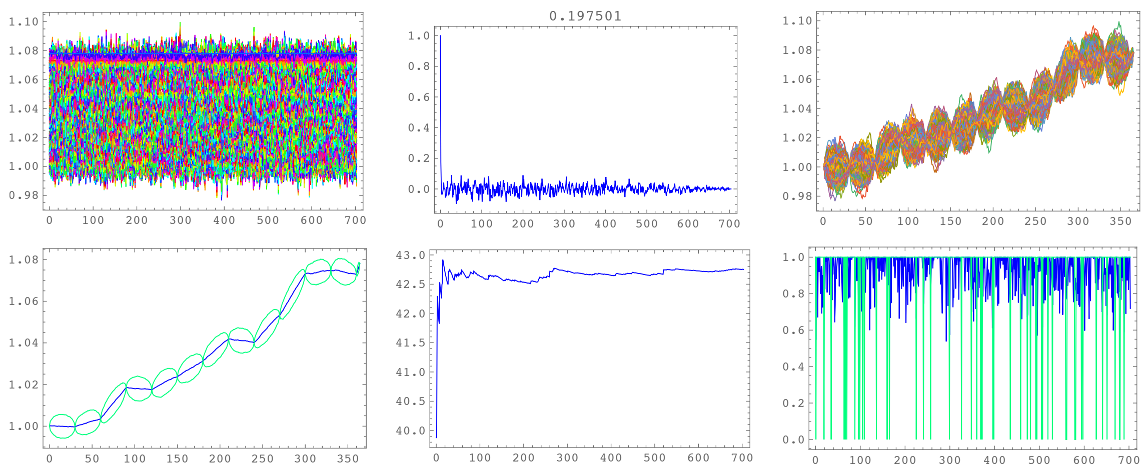

Figure 8: GBM: Langevin sampler, $\hat{p}_2 = \prod_j \phi(\eta_{j+1}, \eta_j|z)/\phi(\eta_j|z)$. From top, left: trajectory $\eta_j(u_l)$ over u_l, autocorrelation of $\eta_J(u_l)$, trajectory $\eta_j(u_l)$ over j, mean $\bar{\eta}_j = L^{-1}\sum_l \eta_j(u_l)$ $\pm$ standard deviation (smoothed trajectory), convergence of estimator $\hat{p}(z)$ (eqn. 13), acceptance probability and rejection indicator for Metropolis algorithm.

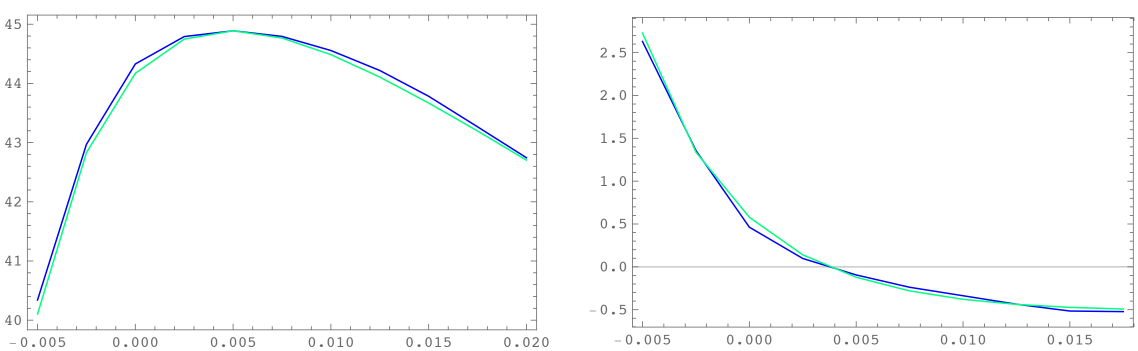

Figure 9: GBM: likelihood (left) and score (right) as a function of $\sigma - 0.02$, $\hat{p}_2 =$ conditional kernel density. Green lines: exact solution.

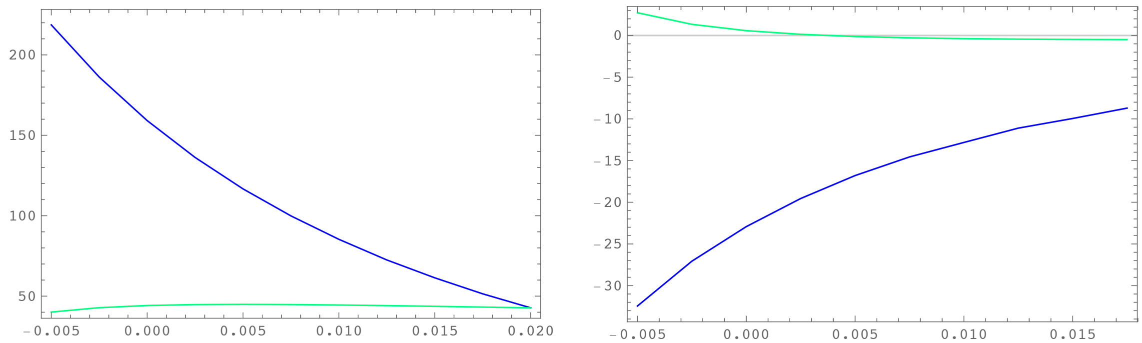

Figure 10: GBM: likelihood and score, $\hat{p}_2$ = full kernel density.

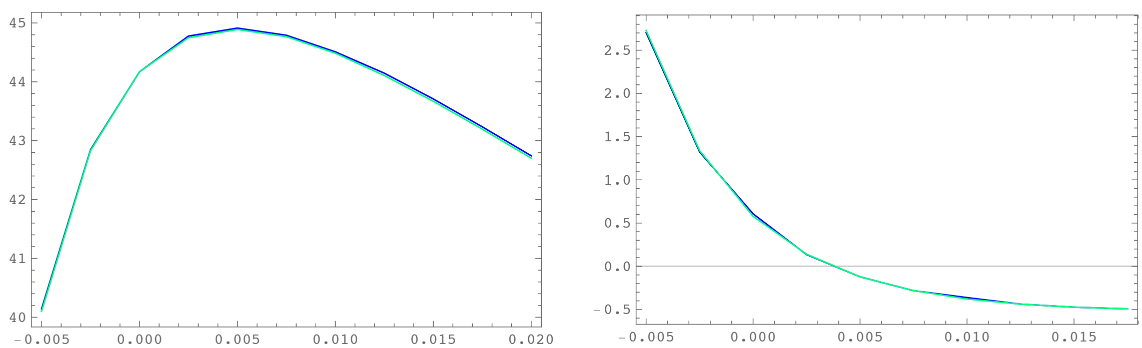

Figure 11: GBM: likelihood and score, $\hat{p}_2$ = conditionally Gaussian transition density.

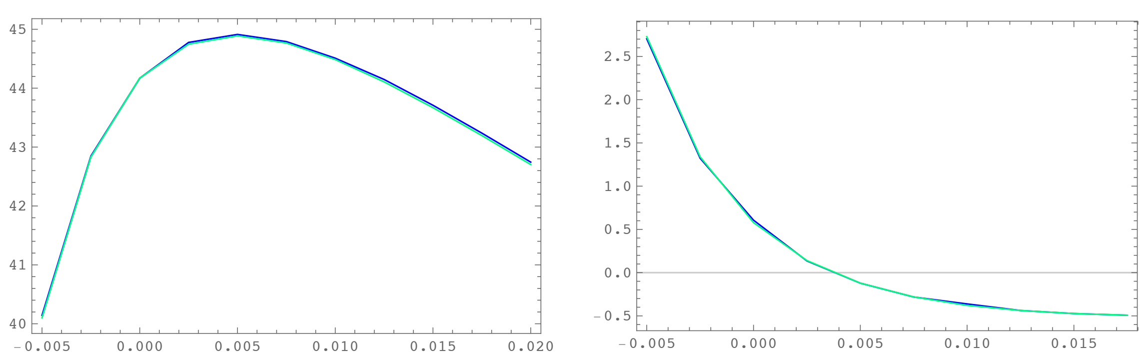

Figure 12: GBM: likelihood and score, $\hat{p}_2$ = linear GLS estimation of drift and diffusion correction $\delta f_j, \delta\Omega_j$ (eqn. 16).

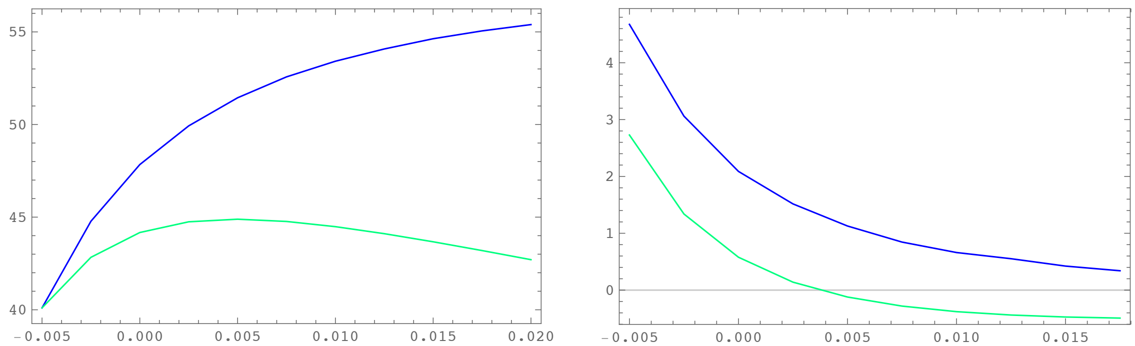

Figure 13: GBM: likelihood and score, $\hat{p}_2$ = linear GLS, constant diffusion matrix.

5.2 Cameron-Martin formula

The functional of the Wiener process

$$E\left[e^{-\frac{\lambda^2}{2}\int_0^T W(t)^2 dt}\right] = 1/\sqrt{\cosh(T\lambda)} \tag{18}$$

was computed analytically by Cameron and Martin (1945); Gelfand and Yaglom (1960). It contains infinitely many coordinates $W(t), 0 \leq t \leq T$. Here a numerical solution is compared with the exact formula (fig. 15). Instead of the output function $p(z|y)$ in eqn. (7) for the likelihood simulation, we use the functional $H = e^{-\frac{\lambda^2}{2}\int_0^T W(t)^2 dt}$, but the importance sampling method applies in the same way. A discretized version of the functional is $H = e^{-\frac{\lambda^2}{2}\sum_{j=0}^{T/\delta t-1} W(t_j)^2 \delta t}$, $\eta_j = W(t_j)$. Clearly, $p_{2,optimal}(\eta) = H(\eta)p(\eta)/\int H(\eta)p(\eta)d\eta$ is a probability density, but not a conditional density.

The output of the Langevin sampler is shown in fig. 14. Since

$$\log H = -\frac{\lambda^2}{2}\sum_{j=0}^{T/\delta t-1} \eta_j^2\, \delta t, \tag{19}$$

and $W(t)$ is Gaussian, we have a quadratic potential Φ and a linear force $-\partial_\eta \Phi$ in the Langevin equation (10). This leads to an acceptance rate of $\alpha = 1$, since we use an Ozaki-type integration method (Ozaki; 1985)(fig. 14, bottom, right).

In fig. 15, the expectation value is shown as function of T. One gets an estimate with very low variance in a small number of replications L. If Ω is not corrected (eqn. 16), the sampling is biased again (fig. 15, right). The relative simulation error is about 1%.

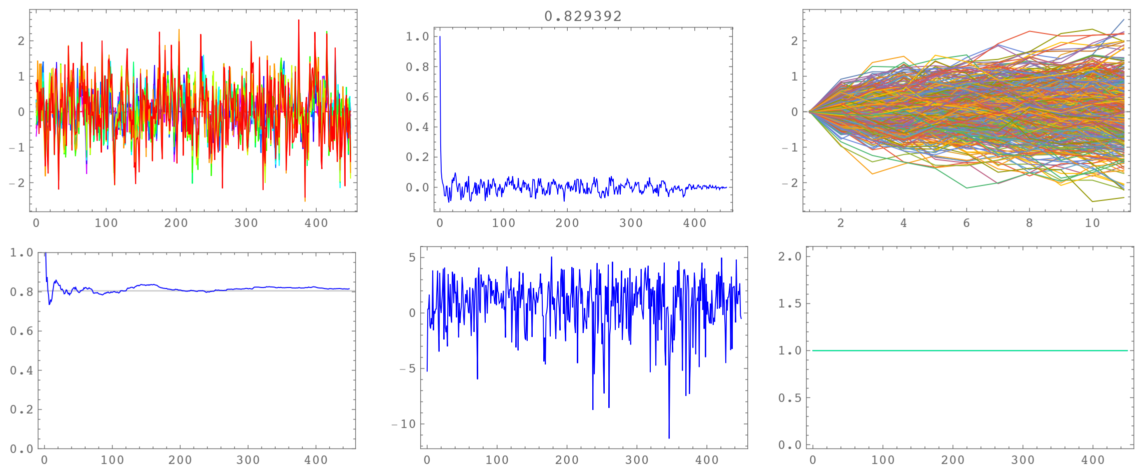

Figure 14: Cameron-Martin formula.
Simulation using a conditionally gaussian importance density. $T = 1, \lambda = 1, dt = 0.1$ and $L = 500$ replications. Exact value $1/\sqrt{\cosh(1)} = 0.805018$. From top left: (11): trajectories $\eta(u_l), l = 0, ..., L$, (12) autocorrelation of $\eta_J(u_l)$, (13) trajectories $\eta_j(u_l), j = 0, ..., J$, bottom left: (21) Convergence of estimate $\hat{H}(u_l)$ over u_l, (22) $\log p_2(u_l)$, (23) Acceptance probability $\alpha(u_l)$.

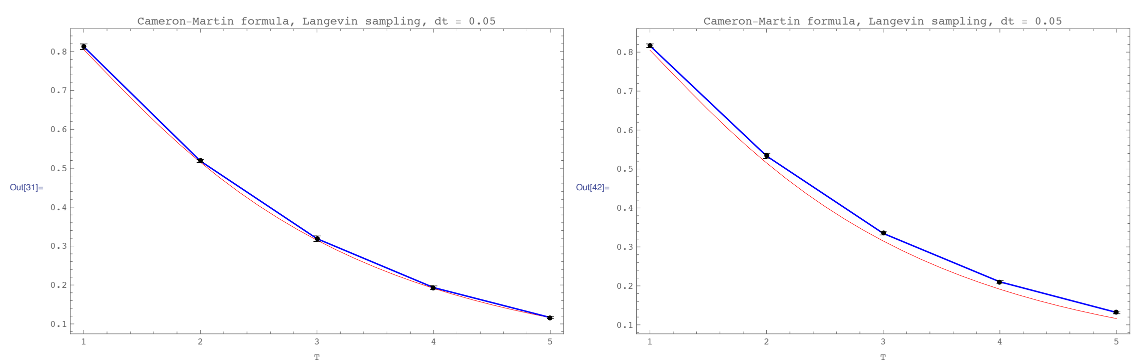

Figure 15: Expectation value as a function of T. Right: $\Omega =$ fix and biased estimates.

5.3 Feynman-Kac-formula

The Schrödinger equation (in imaginary time)[2]

$$u_t = \tfrac{1}{2}u_{xx} - \phi(x)u,$$

with initial condition $u(x, t = 0) = \delta(x - z)$ and quadratic potential (linear oscillator)

$$\phi(x) = \tfrac{1}{2}\gamma^2 x^2$$

can be solved by the Feynman-Kac-formula

$$u(x,t) = E_x\left[e^{-\frac{\gamma^2}{2}\int_0^t W(u)^2 du}\,\delta(W(t) - z)\right] = \sqrt{\frac{\gamma}{2\pi\sinh(\gamma t)}}\exp\left(\frac{\gamma}{2\sinh(\gamma t)}\left[2xz - (x^2 + z^2)\cosh(\gamma t)\right]\right) \tag{20}$$

(Borodin and Salminen; 2002; Feynman and Hibbs; 1965). Here E_x is a conditional expectation value with $W(0) = x$. More generally, one can include a drift term $f(x)u_x$ (see, e.g. Singer; 2014). This describes systems with magnetic fields (see, e.g. Gelfand and Yaglom; 1960) and option pricing in finance (Black and Scholes; 1973; Cox and Ross; 1976). Moreover, setting $\phi = 0$, one can compute the transition density $p(z, t|x, 0)$. Again, the Langevin sampler yields very accurate estimates (fig. 17). Importance sampling is accomplished by simulating trajectories passing through $W(t) = z$ (fig. 16, first row, right picture).

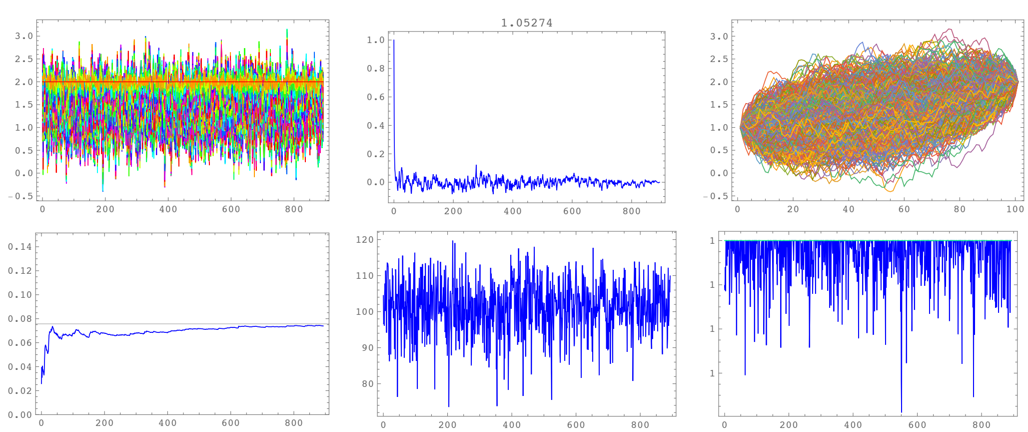

Figure 16: Langevin sampler for the Feynman-Kac-formula.

[2]actually, one has $iu_\tau = -\frac{1}{2}u_{xx} + \phi(x)u; t = i\tau$

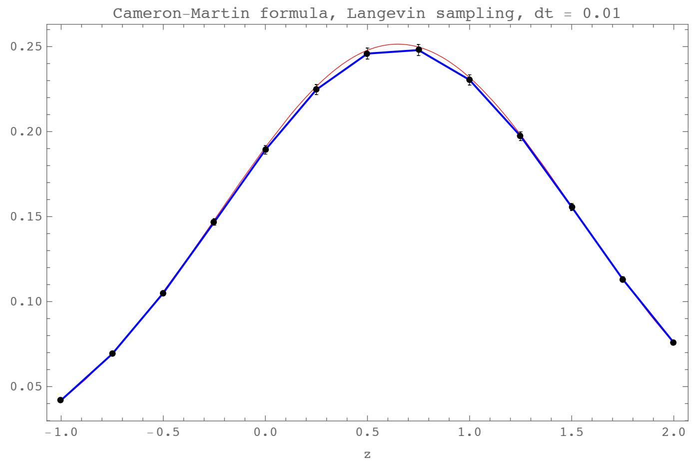

Figure 17: Expectation value as a function of z, $x = 1$, $t = 1$, $\gamma = 1$.

6 Conclusion

Using a Langevin sampler combined with an estimated importance density we are able to compute

- a smooth (w.r.t. parameters) likelihood simulation for nonlinear continuous-discrete state space models.

- perform nonlinear smoothing of latent variables between measurements.

- perform variance reduced MC estimation of functional integrals in finance, statistics and quantum theory (Feynman-Kac formula).

The insertion of latent variables has the disadvantage of producing a high dimensional latent state. The computational burden may be lowered by using improved approximate transition densities, e.g. using the local linearization method (Shoji and Ozaki; 1998a; Singer; 2002), the backward operator method of Aït-Sahalia (2002, 2008); Stramer et al. (2010) or the delta expansion of Li (2013).

Further research will also test other nonlinear models such as the Ginzburg-Landau and the Lorenz model.

References

Aït-Sahalia, Y. (2002). Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach, *Econometrica* **70,1**: 223–262.

Aït-Sahalia, Y. (2008). Closed-form likelihood expansions for multivariate diffusions, *Annals of Statistics* **36, 2**: 906–937.

Bergstrom, A. (1988). The history of continuous-time econometric models, *Econometric Theory* **4**: 365–383.

Bergstrom, A. (ed.) (1976). *Statistical Inference in Continuous Time Models*, North Holland, Amsterdam.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–654.

Borodin, A. and Salminen, P. (2002). *Handbook of Brownian Motion – Facts and Formulae*, second edn, Birkhäuser-Verlag, Basel.

Cameron, R. H. and Martin, W. T. (1945). Transformations of Wiener Integrals Under a General Class of Linear Transformations, *Transactions of the American Mathematical Society* **58**(2): 184–219.

Coleman, J. (1968). The mathematical study of change, *in* H. Blalock and B. A.B. (eds), *Methodology in Social Research*, McGraw-Hill, New York, pp. 428–478.

Cox, J. and Ross, S. (1976). The valuation of options for alternative stochastic processes, *Journal of Financial Economics* **3**: 145–166.

Dargatz, C. (2010). *Bayesian inference for diffusion processes with applications in life sciences*, PhD thesis, LMU, Munich.

Doreian, P. and Hummon, N. (1976). *Modelling Social Processes*, Elsevier, New York, Oxford, Amsterdam.

Doreian, P. and Hummon, N. (1979). Estimating differential equation models on time series: Some simulation evidence, *Sociological Methods and Research* **8**: 3–33.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for simulated maximum likelihood estimation of stochastic differential equations, *Journal of Business and Economic Statistics* **20**: 297–316.

Feynman, R. and Hibbs, A. (1965). *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York.

Gelfand, I. and Yaglom, A. (1960). Integration in Functional Spaces and its Application in Quantum Physics, *Journal of Mathematical Physics* **1**(1): 48–69.

Hairer, M., Stuart, A. M. and Voss, J. (2007). Analysis of SPDEs arising in path sampling, part II: The nonlinear case, *Annals of Applied Probability* **17**(5): 1657–1706.

Hamerle, A., Nagl, W. and Singer, H. (1991). Problems with the estimation of stochastic differential equations using structural equations models, *Journal of Mathematical Sociology* **16, 3**: 201–220.

Jazwinski, A. (1970). *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Kloeden, P. and Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Springer, Berlin. corrected third printing.

Langevin, P. (1908). Sur la théorie du mouvement brownien [on the theory of brownian motion], *Comptes Rendus de l'Academie des Sciences (Paris)* **146**: 530–533.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions, *The Annals of Statistics* **41**(3): 1350–1380.

Nelson, E. (1967). *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton.

Oud, J. and Singer, H. (2008). Special issue: Continuous time modeling of panel data. Editorial introduction, *Statistica Neerlandica* **62, 1**: 1–3.

Ozaki, T. (1985). Nonlinear Time Series and Dynamical Systems, *in* E. Hannan (ed.), *Handbook of Statistics*, North Holland, Amsterdam, pp. 25 – 83.

Pitt, M. (2002). Smooth Particle Filters for Likelihood Evaluation and Maximisation, *Warwick economic research papers 651*, University of Warwick. http://wrap.warwick.ac.uk/1536/.

Reznikoff, M. and Vanden-Eijnden, E. (2005). Invariant measures of stochastic partial differential equations and conditioned diffusions, *C. R. Acad. Sci. Paris* **Ser. I 340**: 305–308.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the metropolis–hastings algorithm, *Biometrika* **88**(3): 603–621.

Roberts, G. O. and Stramer, O. (2002). Langevin Diffusions and Metropolis-Hastings Algorithms, *Methodology And Computing In Applied Probability* **4**(4): 337–357.

Schweppe, F. (1965). Evaluation of likelihood functions for gaussian signals, *IEEE Transactions on Information Theory* **11**: 61–70.

Shoji, I. and Ozaki, T. (1998a). A statistical method of estimation and simulation for systems of stochastic differential equations, *Biometrika* **85, 1**: 240–243.

Shoji, I. and Ozaki, T. (1998b). Estimation for nonlinear stochastic differential equations by a local linearization method 1, *Stochastic Analysis and Applications* **16**(4): 733–752.

Simon, H. (1952). A formal theory of interaction in social groups, *American Sociological Review* **17**: 202–211.

Singer, H. (1990). *Parameterschätzung in zeitkontinuierlichen dynamischen Systemen [Parameter estimation in continuous time dynamical systems; Ph.D. thesis, University of Konstanz, in German]*, Hartung-Gorre-Verlag, Konstanz.

Singer, H. (1995). Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values, *Econometric Theory* **11**: 721–735.

Singer, H. (2002). Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood vs. Extended Kalman Filter and Itô-Taylor Expansion, *Journal of Computational and Graphical Statistics* **11**(4): 972–995.

Singer, H. (2003). Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing, *Computational Statistics* **18**(1): 79–106.

Singer, H. (2007). Stochastic Differential Equation Models with Sampled Data, *in* K. van Montfort, J. Oud and A. Satorra (eds), *Longitudinal Models in the Behavioral and Related Sciences*, The European Association of Methodology (EAM) Methodology and Statistics series, vol. II, Lawrence Erlbaum Associates, Mahwah, London, pp. 73–106.

Singer, H. (2011). Continuous-discrete state-space modeling of panel data with nonlinear filter algorithms, *Advances in Statistical Analysis* **95**: 375–413.

Singer, H. (2012). SEM modeling with singular moment matrices. Part II: ML-Estimation of Sampled Stochastic Differential Equations., *Journal of Mathematical Sociology* **36**(1): 22–43.

Singer, H. (2014). Importance Sampling for Kolmogorov Backward Equations, *Advances in Statistical Analysis* **98**: 345–369.

Singer, H. (2015). Conditional Gauss–Hermite filtering with application to volatility estimation, *IEEE Transactions on Automatic Control* **60**(9): 2476–2481.

Singer, H. (2016). Maximum Likelihood Estimation of Continuous-Discrete State-Space Models: Langevin Path Sampling vs. Numerical Integration, *Diskussionsbeiträge Fakultät Wirtschaftswissenschaft*, FernUniversität in Hagen. http://www.fernuni-hagen.de/wirtschaftswissenschaft/forschung/beitraege.shtml.

Stramer, O., Bognar, M. and Schneider, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions, *Journal of Financial Econometrics* **8**(4): 450–480.

Tanner, M. (1996). *Tools for Statistical Inference*, third edn, Springer, New York.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation, *Journal of the American statistical Association* **82**(398): 528–540.

Die Diskussionspapiere ab Nr. 183 (1992) bis heute, können Sie im Internet unter
http://www.fernuni-hagen.de/wirtschaftswissenschaft/forschung/beitraege.shtml einsehen
und zum Teil downloaden.
Ältere Diskussionspapiere selber erhalten Sie nur in den Bibliotheken.

Nr	Jahr	Titel	Autor/en
420	2008	Stockkeeping and controlling under game theoretic aspects	Fandel, Günter Trockel, Jan
421	2008	On Overdissipation of Rents in Contests with Endogenous Intrinsic Motivation	Schleptütz, Volker
422	2008	Maximum Entropy Inference for Mixed Continuous-Discrete Variables	Singer, Hermann
423	2008	Eine Heuristik für das mehrdimensionale Bin Packing Problem	Mack, Daniel Bortfeldt, Andreas
424	2008	Expected A Posteriori Estimation in Financial Applications	Mazzoni, Thomas
425	2008	A Genetic Algorithm for the Two-Dimensional Knapsack Problem with Rectangular Pieces	Bortfeldt, Andreas Winter, Tobias
426	2008	A Tree Search Algorithm for Solving the Container Loading Problem	Fanslau, Tobias Bortfeldt, Andreas
427	2008	Dynamic Effects of Offshoring	Stijepic, Denis Wagner, Helmut
428	2008	Der Einfluss von Kostenabweichungen auf das Nash-Gleichgewicht in einem nicht-kooperativen Disponenten-Controller-Spiel	Fandel, Günter Trockel, Jan
429	2008	Fast Analytic Option Valuation with GARCH	Mazzoni, Thomas
430	2008	Conditional Gauss-Hermite Filtering with Application to Volatility Estimation	Singer, Hermann
431	2008	Web 2.0 auf dem Prüfstand: Zur Bewertung von Internet-Unternehmen	Christian Maaß Gotthard Pietsch
432	2008	Zentralbank-Kommunikation und Finanzstabilität – Eine Bestandsaufnahme	Knütter, Rolf Mohr, Benjamin
433	2008	Globalization and Asset Prices: Which Trade-Offs Do Central Banks Face in Small Open Economies?	Knütter, Rolf Wagner, Helmut
434	2008	International Policy Coordination and Simple Monetary Policy Rules	Berger, Wolfram Wagner, Helmut
435	2009	Matchingprozesse auf beruflichen Teilarbeitsmärkten	Stops, Michael Mazzoni, Thomas
436	2009	Wayfindingprozesse in Parksituationen - eine empirische Analyse	Fließ, Sabine Tetzner, Stefan
437	2009	ENTROPY-DRIVEN PORTFOLIO SELECTION a downside and upside risk framework	Rödder, Wilhelm Gartner, Ivan Ricardo Rudolph, Sandra
438	2009	Consulting Incentives in Contests	Schleptütz, Volker

439	2009	A Genetic Algorithm for a Bi-Objective Winner-Determination Problem in a Transportation-Procurement Auction"	Buer, Tobias Pankratz, Giselher
440	2009	Parallel greedy algorithms for packing unequal spheres into a cuboidal strip or a cuboid	Kubach, Timo Bortfeldt, Andreas Tilli, Thomas Gehring, Hermann
441	2009	SEM modeling with singular moment matrices Part I: ML-Estimation of time series	Singer, Hermann
442	2009	SEM modeling with singular moment matrices Part II: ML-Estimation of sampled stochastic differential equations	Singer, Hermann
443	2009	Konsensuale Effizienzbewertung und -verbesserung – Untersuchungen mittels der Data Envelopment Analysis (DEA)	Rödder, Wilhelm Reucher, Elmar
444	2009	Legal Uncertainty – Is Hamonization of Law the Right Answer? A Short Overview	Wagner, Helmut
445	2009	Fast Continuous-Discrete DAF-Filters	Mazzoni, Thomas
446	2010	Quantitative Evaluierung von Multi-Level Marketingsystemen	Lorenz, Marina Mazzoni, Thomas
447	2010	Quasi-Continuous Maximum Entropy Distribution Approximation with Kernel Density	Mazzoni, Thomas Reucher, Elmar
448	2010	Solving a Bi-Objective Winner Determination Problem in a Transportation Procurement Auction	Buer, Tobias Pankratz, Giselher
449	2010	Are Short Term Stock Asset Returns Predictable? An Extended Empirical Analysis	Mazzoni, Thomas
450	2010	Europäische Gesundheitssysteme im Vergleich – Effizienzmessungen von Akutkrankenhäusern mit DEA –	Reucher, Elmar Sartorius, Frank
451	2010	Patterns in Object-Oriented Analysis	Blaimer, Nicolas Bortfeldt, Andreas Pankratz, Giselher
452	2010	The Kuznets-Kaldor-Puzzle and Neutral Cross-Capital-Intensity Structural Change	Stijepic, Denis Wagner, Helmut
453	2010	Monetary Policy and Boom-Bust Cycles: The Role of Communication	Knütter, Rolf Wagner, Helmut
454	2010	Konsensuale Effizienzbewertung und –verbesserung mittels DEA – Output- vs. Inputorientierung –	Reucher, Elmar Rödder, Wilhelm
455	2010	Consistent Modeling of Risk Averse Behavior with Spectral Risk Measures	Wächter, Hans Peter Mazzoni, Thomas

456	2010	Der virtuelle Peer – Eine Anwendung der DEA zur konsensualen Effizienz-bewertung –	Reucher, Elmar
457	2010	A two-stage packing procedure for a Portuguese trading company	Moura, Ana Bortfeldt, Andreas
458	2010	A tree search algorithm for solving the multi-dimensional strip packing problem with guillotine cutting constraint	Bortfeldt, Andreas Jungmann, Sabine
459	2010	Equity and Efficiency in Regional Public Good Supply with Imperfect Labour Mobility – Horizontal versus Vertical Equalization	Arnold, Volker
460	2010	A hybrid algorithm for the capacitated vehicle routing problem with three-dimensional loading constraints	Bortfeldt, Andreas
461	2010	A tree search procedure for the container relocation problem	Forster, Florian Bortfeldt, Andreas
462	2011	Advanced X-Efficiencies for CCR- and BCC-Modell – Towards Peer-based DEA Controlling	Rödder, Wilhelm Reucher, Elmar
463	2011	The Effects of Central Bank Communication on Financial Stability: A Systematization of the Empirical Evidence	Knütter, Rolf Mohr, Benjamin Wagner, Helmut
464	2011	Lösungskonzepte zur Allokation von Kooperationsvorteilen in der kooperativen Transportdisposition	Strangmeier, Reinhard Fiedler, Matthias
465	2011	Grenzen einer Legitimation staatlicher Maßnahmen gegenüber Kreditinstituten zur Verhinderung von Banken- und Wirtschaftskrisen	Merbecks, Ute
466	2011	Controlling im Stadtmarketing – Eine Analyse des Hagener Schaufensterwettbewerbs 2010	Fließ, Sabine Bauer, Katharina
467	2011	A Structural Approach to Financial Stability: On the Beneficial Role of Regulatory Governance	Mohr, Benjamin Wagner, Helmut
468	2011	Data Envelopment Analysis - Skalenerträge und Kreuzskalenerträge	Wilhelm Rödder Andreas Dellnitz
469	2011	Controlling organisatorischer Entscheidungen: Konzeptionelle Überlegungen	Lindner, Florian Scherm, Ewald
470	2011	Orientierung in Dienstleistungsumgebungen – eine explorative Studie am Beispiel des Flughafen Frankfurt am Main	Fließ, Sabine Colaci, Antje Nesper, Jens

471	2011	Inequality aversion, income skewness and the theory of the welfare state	Weinreich, Daniel
472	2011	A tree search procedure for the container retrieval problem	Forster, Florian Bortfeldt, Andreas
473	2011	A Functional Approach to Pricing Complex Barrier Options	Mazzoni, Thomas
474	2011	Bologna-Prozess und neues Steuerungsmodell – auf Konfrontationskurs mit universitären Identitäten	Jost, Tobias Scherm, Ewald
475	2011	A reduction approach for solving the rectangle packing area minimization problem	Bortfeldt, Andreas
476	2011	Trade and Unemployment with Heterogeneous Firms: How Good Jobs Are Lost	Altenburg, Lutz
477	2012	Structural Change Patterns and Development: China in Comparison	Wagner, Helmut
478	2012	Demografische Risiken – Herausforderungen für das finanzwirtschaftliche Risikomanagement im Rahmen der betrieblichen Altersversorgung	Merbecks, Ute
479	2012	"It's all in the Mix!" – Internalizing Externalities with R&D Subsidies and Environmental Liability	Endres, Alfred Friehe, Tim Rundshagen, Bianca
480	2012	Ökonomische Interpretationen der Skalenvariablen u in der DEA	Dellnitz, Andreas Kleine, Andreas Rödder, Wilhelm
481	2012	Entropiebasierte Analyse von Interaktionen in Sozialen Netzwerken	Rödder, Wilhelm Brenner, Dominic Kulmann, Friedhelm
482	2013	Central Bank Independence and Financial Stability: A Tale of Perfect Harmony?	Berger, Wolfram Kißmer, Friedrich
483	2013	Energy generation with Directed Technical Change	Kollenbach, Gilbert
484	2013	Monetary Policy and Asset Prices: When Cleaning Up Hits the Zero Lower Bound	Berger, Wolfram Kißmer, Friedrich
485	2013	Superknoten in Sozialen Netzwerken – eine entropieoptimale Analyse	Brenner, Dominic, Rödder, Wilhelm, Kulmann, Friedhelm
486	2013	Stimmigkeit von Situation, Organisation und Person: Gestaltungsüberlegungen auf Basis des Informationsverarbeitungsansatzes	Julmi, Christian Lindner, Florian Scherm, Ewald
487	2014	Incentives for Advanced Abatement Technology Under National and International Permit Trading	Endres, Alfred Rundshagen, Bianca

488	2014	Dynamische Effizienzbewertung öffentlicher Dreispartentheater mit der Data Envelopment Analysis	Kleine, Andreas Hoffmann, Steffen
489	2015	Konsensuale Peer-Wahl in der DEA -- Effizienz vs. Skalenertrag	Dellnitz, Andreas Reucher, Elmar
490	2015	Makroprudenzielle Regulierung – eine kurze Einführung und ein Überblick	Velauthapillai, Jeyakrishna
491	2015	SEM modeling with singular moment matrices Part III: GLS estimation	Singer, Hermann
492	2015	Die steuerliche Berücksichtigung von Aufwendungen für ein Studium − Eine Darstellung unter besonderer Berücksichtigung des Hörerstatus	Meyering, Stephan Portheine, Kea
493	2016	Ungewissheit versus Unsicherheit in Sozialen Netzwerken	Rödder, Wilhelm Dellnitz, Andreas Gartner, Ivan
494	2016	Investments in supplier-specific economies of scope with two different services and different supplier characters: two specialists	Fandel, Günter Trockel, Jan
495	2016	An application of the put-call-parity to variance reduced Monte-Carlo option pricing	Müller, Armin
496	2016	A joint application of the put-call-parity and importance sampling to variance reduced option pricing	Müller, Armin
497	2016	Simulated Maximum Likelihood for Continuous-Discrete State Space Models using Langevin Importance Sampling	Singer, Hermann