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

Hermann Singer

FernUniversität in Hagen,
Hermann.Singer@FernUni-Hagen.de

presented at the

9th International Conference on Social Science Methodology (RC33)

11.-16. September 2016, Leicester, UK

Session:
Estimation of Stochastic Differential Equations with Time Series, Panel and Spatial Data.

Nonlinear continuous/discrete state space model

$$dY(t) = f(Y(t), t)dt + g(Y(t), t)dW(t)$$

$$Z_i = h(Y_i, t_i) + \epsilon_i$$

$$i = 0, ..., T$$

- nonlinear drift and diffusion functions f, g
  $f = f(Y(t), t, x(t), \psi)$
- nonlinear diffusion $g(Y)$: Itô calculus
- Spatial models: $Y_n(t) = Y(x_n, t), x_n \in \mathbb{R}^d$:
  Random field $Y(x, t, \omega)$

Linear stochastic differential equations (LSDE)

(exact ML, Singer; 1990, Kalman filter)

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

- $W(t, \omega)$: Wiener process:
  continuous time random walk (Brownian motion)
- Itô stochastic differential equations:
  $dW/dt = \zeta(t) =$ Gaussian white noise
- A: drift matrix
- G: diffusion matrix

Wiener process and stock index

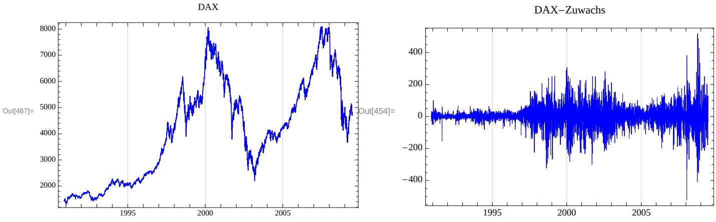

Figure : German stock index (DAX)

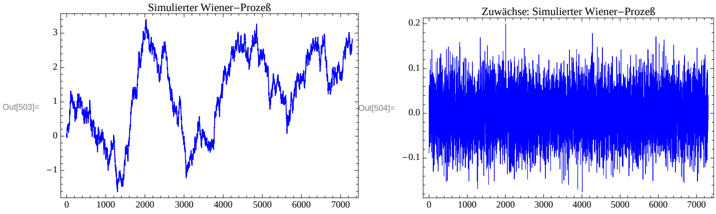

Figure : Simulated Wiener process (random walk)

Simulated Wiener processes

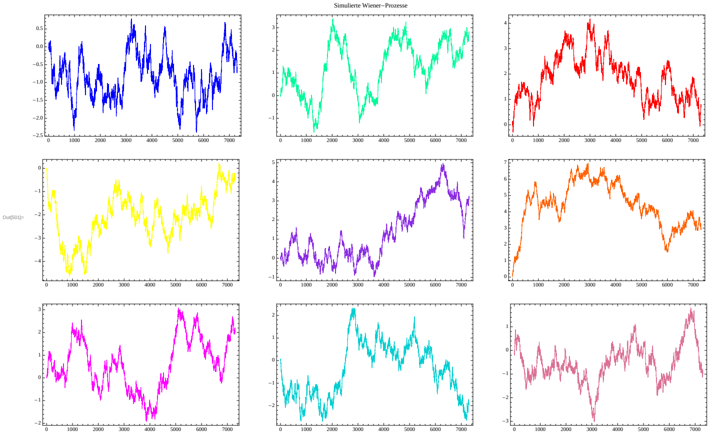

Figure : Simulated Wiener processes

Exact discrete model (EDM)

Bergstrom (1976, 1988)

$$Y_{i+1} = e^{A(t_{i+1}-t_i)} Y_i + \int_{t_i}^{t_{i+1}} e^{A(t_{i+1}-s)} G dW(s)$$

restricted VAR(1) model

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

- Φ: fundamental matrix of the system
- $Y_i := Y(t_i)$: sampled measurements

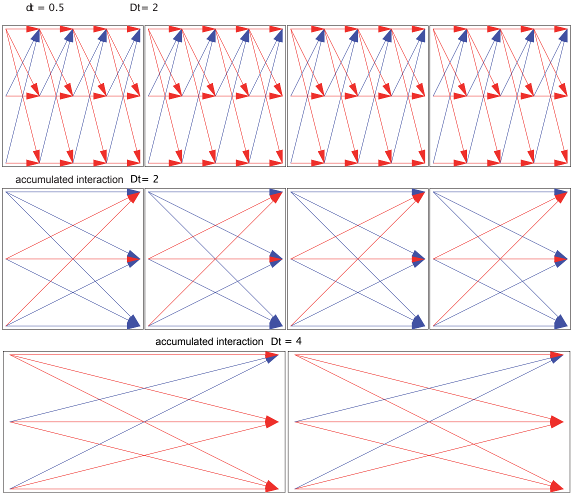

Figure : 3 variable model:
Product representation of matrix exponential within the measurement interval $\Delta t = 2$. Latent states η_j, discretization interval $\delta t = 2/4 = 0.5$ (Singer; 2012)

Maximum Likelihood Parameter Estimation

likelihood function of all observations

$$p(z_T, ..., z_0) \quad = \quad \int p(z_T, ..., z_0 | y_T, ..., y_0) p(y_T, ..., y_0) dy$$

- High dimensional integration over latent variables
- smooth dependence on parameter vector ψ:
  $L(\psi) = p(z; \psi)$
- Problem: $p(y_T, ..., y_0)$ not known
  Sampling interval Δt_i
  Transition density $p(y_{i+1}, \Delta t_i | y_i)$ difficult to compute
- Use additional latent variables
  $y_T = \eta_J, ..., \eta_0 = y_0$, $y(t_i) = y_i = \eta_{j_i}$

Integration

latent states η_j

$$p(z_T,...,z_0) = \int p(z_T,...,z_0|\eta_J,...,\eta_0)p(\eta_J,...,\eta_0)d\eta$$
$$= E\Big[p(z_T,...,z_0|\eta_J,...,\eta_0)\Big]$$

- $\eta_{j_i} = y(t_i) = y_i,\ t_i =$ measurement times
- even higher (infinite) dimensional integration over latent variables:
  Markov Chain Monte Carlo: simulate η
- Euler density (discretization interval δt)

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

$$f_j = f(\eta_j, \tau_j), \Omega_j = (gg')(\eta_j, \tau_j)$$

Importance Sampling

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

- p_2: importance density
- $p_{2,optimal} = \frac{p(z|\eta)p(\eta)}{p(z)} = p(\eta|z)$
- however, $p(z)$ is unknown
- $\eta \to \eta(t, u)$: random field, $u =$ simulation time

Langevin Sampling

Langevin (1908); Roberts and Stramer (2002)

Langevin equation

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

- stationary distribution of conditional latent states

$$p_{stat}(\eta) = e^{-\Phi(\eta)} = p(\eta|z)$$

- drift = −gradient of potential $\Phi(\eta)$

$$-\partial_\eta \Phi(\eta) = \partial_\eta[\log p(z|\eta) + \log p(\eta) - \log p(z)]$$

  sample from $p(\eta|z)$, $p(z)$ not needed!

- optimal nonlinear smoothing
  $\eta(u) \sim p(\eta|z)$ in equilibrium $u \to \infty$

Simulated Likelihood

Variance reduced MC-integration

$$\hat{p}(z_T, ..., z_0) = L^{-1} \sum p(z|\eta_l) \frac{p(\eta_l)}{p_2(\eta_l)}$$

- $\eta_l \sim p(\eta|z)$ in equilibrium
- draw optimal $\eta_l = \eta(u_l)$ from discretized Langevin equation (including Metropolis-Hastings mechanism)
- $p_{2,optimal} = \frac{p(z|\eta)p(\eta)}{p(z)} = p(\eta|z)$ is unknown
- Idea: estimate $p_2 = p(\eta|z)$ from $\eta_l \sim p(\eta|z)$

Estimation of importance density

- use known (suboptimal) reference density

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

- kernel density estimate

$$\hat{p}(\eta|z) = L^{-1}\sum_l k(\eta - \eta_l; smooth)$$

Problem:

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

Estimation of importance density

- use Markov structure of state space model

$$\begin{aligned} \eta_{j+1} &= f(\eta_j)\delta t + g(\eta_j)\delta W_j \\ z_i &= h(y_i) + \epsilon_i \\ p(\eta|z) &= p(\eta_J|\eta_{J-1}, ..., \eta_0, z) \, * p(\eta_{J-1}, ..., \eta_0|z) \end{aligned}$$

Conditional Markov process

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

use conditional independence of past $z^i = (z_0, ..., z_i)$ and future $\bar{z}^i = (z_{i+1}, ...z_T)$ given η^j:

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

Euler transition kernel

- 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

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

- nonlinear regression for δf_j and $\delta\Omega_j$ (parametric and nonparametric)
- draw data $\eta_{jl} = \eta(\tau_j, u_l) \sim p(\eta|z)$

Kernel density

conditional transition density

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

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

Examples: Geometrical Brownian motion

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

- nonlinear model (multiplicative noise) $y * dW$
- exact solution: 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 discrete model

$$y(t) = y(t_0)e^{(\mu-\sigma^2/2)(t-t_0)+\sigma[W(t)-W(t_0)]}$$

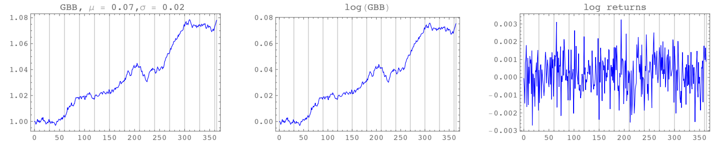

Figure : Trajectory and log returns.

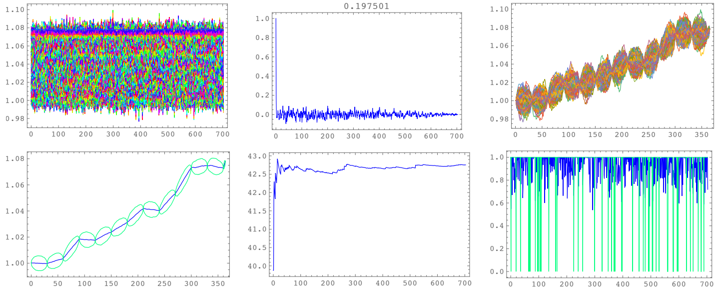

Figure : Langevin sampler, $\hat{p}_2 = \prod_j \phi(\eta_{j+1}, \eta_j | z)/\phi(\eta_j | z)$.

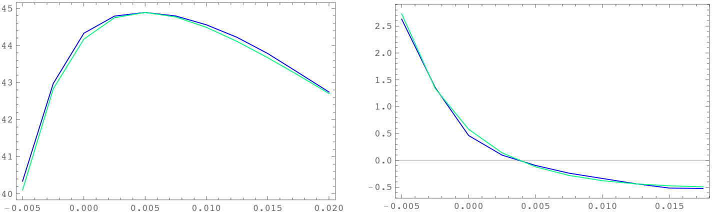

Figure : likelihood and score, $\hat{p}_2$ = conditional kernel density.

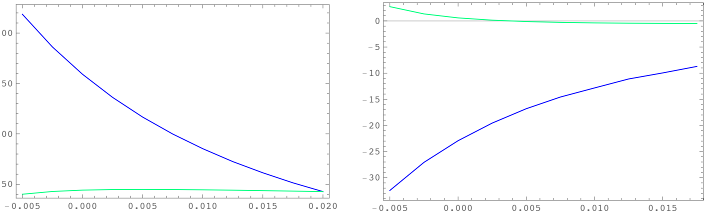

Figure : likelihood and score, $\hat{p}_2$ = full kernel density.

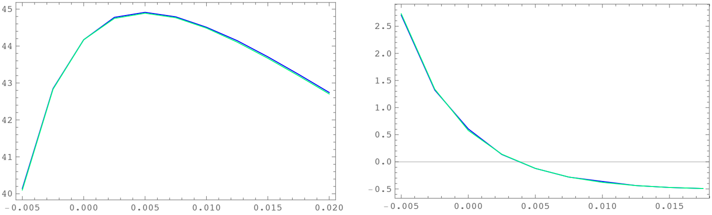

Figure : likelihood and score, $\hat{p}_2$ = conditionally Gaussian.

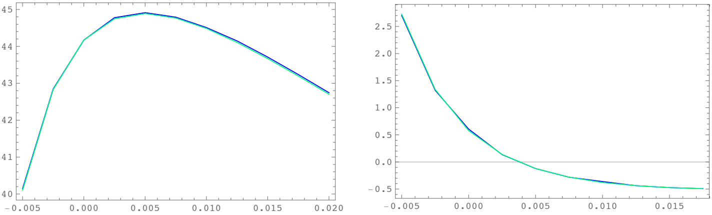

Figure : likelihood and score, $\hat{p}_2$ = linear GLS.

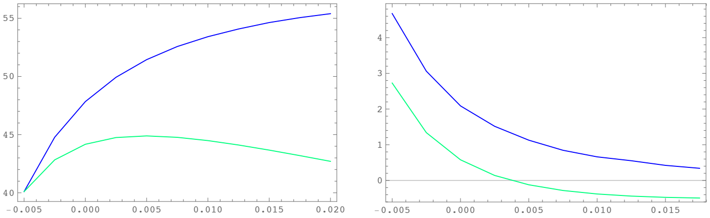

Figure : likelihood and score, $\hat{p}_2 =$ linear GLS, constant diffusion matrix.

Examples: Cameron-Martin formula

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

Cameron and Martin (1945); Gelfand and Yaglom (1960)

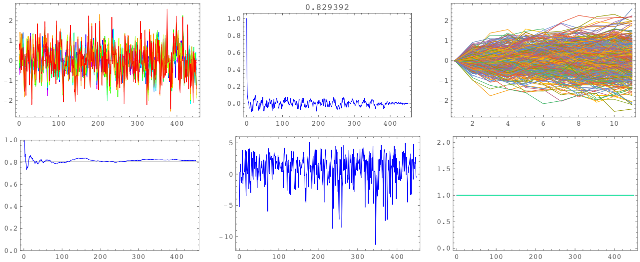

Figure : Cameron-Martin formula.
Simulation using a conditionally gaussian importance density.
$T = 1, \lambda = 1, dt = 0.1$ and $L = 500$ replications. Exact value $1/\sqrt{\cosh(1)} = 0.805018$

Examples: Cameron-Martin formula

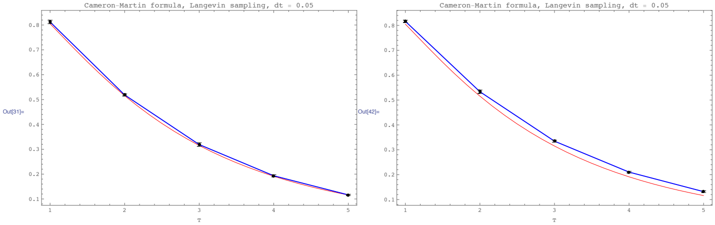

Figure : Expectation value as a function of T. Right: $\Omega = \text{fix}$.

Examples: Feynman-Kac-formula

Schrödinger Equation (imaginary time $t = i\tau$)

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

$$\begin{aligned} u(x, t=0) &= \delta(x - z) \\ \phi(x) &= \frac{1}{2}\gamma^2 x^2 \end{aligned}$$

Feynman-Kac-formula

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

Feynman-Kac-formula

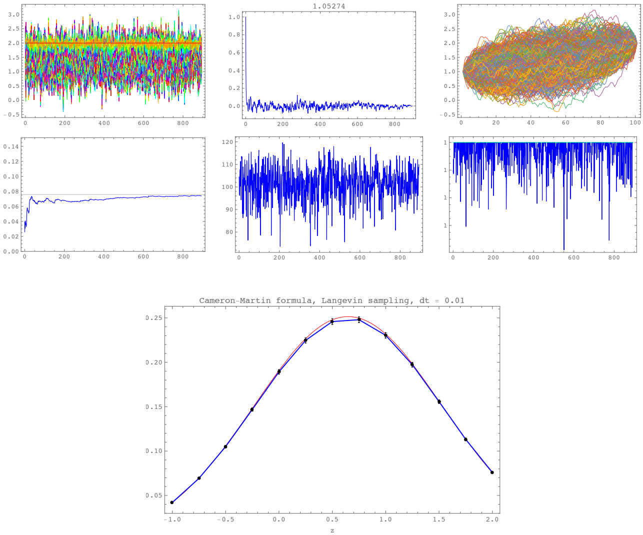

Figure : Expectation value as a function of z, $x = 1$, $t = 1$, $\gamma = 1$.

Conclusion

- smooth likelihood simulation for nonlinear continuous-discrete state space models.
- Nonlinear smoothing of latent variables between measurements.
- Variance reduced MC estimation of functional integrals in finance, statistics and quantum theory (Feynman-Kac formula).

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.

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.

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.

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.

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

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

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