MCMC schemes for partially observed diffusions

Gavin Whitaker

School of Mathematics and Statistics
Newcastle University

26th March 2013

Supervisors: Richard Boys and Andrew Golightly

SDE Models

- Consider an Itô process $\{\boldsymbol{X}_t, t \geq 0\}$ satisfying

$$d\boldsymbol{X}_t = \boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})dt + \boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})^{\frac{1}{2}} d\boldsymbol{W}_t$$

  - $\boldsymbol{X}_t$ is the value of the process at time t
  - $\boldsymbol{\theta}$ is the length p parameter vector
  - $\boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is the drift
  - $\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})$ is the diffusion coefficient
  - $\boldsymbol{W}_t$ is standard Brownian motion
  - $\boldsymbol{X}_0$ is the vector of initial conditions

- Seek a numerical solution via (for example) the Euler-Maruyama approximation

$$\Delta\boldsymbol{X}_t \equiv \boldsymbol{X}_{t+\Delta t} - \boldsymbol{X}_t = \boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})\Delta t + \boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})^{\frac{1}{2}} \Delta\boldsymbol{W}_t$$

where $\Delta\boldsymbol{W}_t \sim N(\boldsymbol{0}, \boldsymbol{I}\Delta t)$

SDE Models

Lotka-Volterra Model

$$
\begin{array}{lccc}
\textbf{Reaction 1:} & \mathcal{X}_1 & \xrightarrow{\theta_1} & 2\mathcal{X}_1 \\
\textbf{Reaction 2:} & \mathcal{X}_1 + \mathcal{X}_2 & \xrightarrow{\theta_2} & 2\mathcal{X}_2 \\
\textbf{Reaction 3:} & \mathcal{X}_2 & \xrightarrow{\theta_3} & \phi
\end{array}
$$

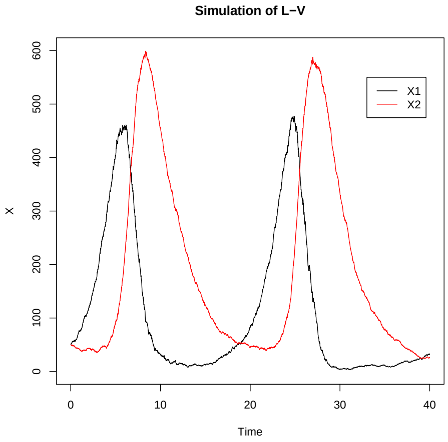
The mass action SDE representation of the system dynamics is

$$
\begin{aligned}
d\boldsymbol{X}_t &= \begin{pmatrix} \theta_1 X_1 - \theta_2 X_1 X_2 \\ \theta_2 X_1 X_2 - \theta_3 X_2 \end{pmatrix} dt \\
&\quad + \begin{pmatrix} \theta_1 X_1 + \theta_2 X_1 X_2 & -\theta_2 X_1 X_2 \\ -\theta_2 X_1 X_2 & \theta_3 X_2 + \theta_2 X_1 X_2 \end{pmatrix}^{\frac{1}{2}} d\boldsymbol{W}_t,
\end{aligned}
$$

After dropping dependence of $\boldsymbol{X}_t$ on t for simplicity

Lotka-Volterra Model

Figure: Numerical solution for L-V model, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Bayesian inference for SDEs

- Problematic due to the intractability of the transition density characterising the process
- In other words, we typically can't analytically solve the SDE

$$d\boldsymbol{X}_t = \boldsymbol{\alpha}(\boldsymbol{X}_t, \boldsymbol{\theta})dt + \boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta})^{\frac{1}{2}} d\boldsymbol{W}_t$$

- So we could just work with the Euler-Maruyama approximation
- Suppose we have data $\boldsymbol{d}$ at equidistant times $t_0, t_1, \ldots, t_n$, where $t_{i+1} - t_i = \Delta t$
- The Euler-Maruyama approximation might not be accurate if Δt is too large
- We therefore adopt a data augmentation approach

Bayesian inference for SDEs

- Consider $[t_i, t_{i+1}]$. Insert $m-1$ additional time points

$$t_i = \tau_{im} < \tau_{im+1} < \ldots < \tau_{(i+1)m} = t_{i+1}$$

  where

$$\Delta\tau \equiv \tau_{im+1} - \tau_{im} = \frac{t_{i+1} - t_i}{m}$$

- We don't know the value of the process at these additional (latent) times
- This is a data augmentation approach
- Apply Euler-Maruyama approximation over each interval of length $\Delta\tau$

Bayesian inference for SDEs

- Formulate joint posterior for parameters and latent values

$$
\begin{aligned}
\rightarrow \quad \boldsymbol{d} &= (\boldsymbol{x}_{t_0}, \boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_n}) \\
\rightarrow \quad \boldsymbol{x} &= (\boldsymbol{x}_{\tau_1}, \boldsymbol{x}_{\tau_2}, \ldots, \boldsymbol{x}_{\tau_{m-1}}, \boldsymbol{x}_{\tau_{m+1}}, \ldots, \ldots, \boldsymbol{x}_{\tau_{nm-1}}) \\
&= \text{latent path} \\
\rightarrow \quad (\boldsymbol{x}, \boldsymbol{d}) &= (\boldsymbol{x}_{\tau_0}, \boldsymbol{x}_{\tau_1}, \ldots, \boldsymbol{x}_{\tau_m}, \boldsymbol{x}_{\tau_{m+1}}, \ldots, \ldots, \boldsymbol{x}_{\tau_{nm}}) \\
&= \text{augmented path}
\end{aligned}
$$

Formulate joint posterior for parameters and latent data as

$$\begin{aligned}\pi(\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{d}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x},\boldsymbol{d}|\boldsymbol{\theta}) \\ &\propto \underbrace{\pi(\boldsymbol{\theta})}_{\text{prior}} \prod_{i=0}^{nm-1} \underbrace{\pi(\boldsymbol{x}_{\tau_{i+1}}|\boldsymbol{x}_{\tau_i},\boldsymbol{\theta})}_{\text{Euler density}} \end{aligned} \tag{1}$$

where

$$\pi(\boldsymbol{x}_{\tau_{i+1}}|\boldsymbol{x}_{\tau_i},\boldsymbol{\theta}) = \phi\left(\boldsymbol{x}_{\tau_{i+1}}\,;\,\boldsymbol{x}_{\tau_i}+\boldsymbol{\alpha}(\boldsymbol{x}_{\tau_i}\,,\,\boldsymbol{\theta})\Delta t,\boldsymbol{\beta}(\boldsymbol{x}_{\tau_i},\boldsymbol{\theta})\Delta t\right)$$

and $\phi(\cdot\,;\,\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the Gaussian density with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$

- ▷ The posterior distribution is typically analytically intractable

A Gibbs sampling approach

- We therefore sample via an MCMC scheme
- E.g a Gibbs sampler, alternating between draws of
  - $\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}$
  - $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}$
- The last step can be done (for example) in blocks of length $m-1$ between observations
- Metropolis within Gibbs updates may be needed

Sampling $\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}$

$$
\begin{aligned}
\pi(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{d}) &\propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta}) \\
&\propto \underbrace{\pi(\boldsymbol{\theta})}_{\text{prior}} \prod_{i=0}^{nm-1} \underbrace{\pi(\boldsymbol{x}_{\tau_{i+1}}|\boldsymbol{x}_{\tau_i}, \boldsymbol{\theta})}_{\text{Euler density}}
\end{aligned}
$$

- Typically intractable so use a M-H step
- Propose

$$\boldsymbol{\theta}^*|\boldsymbol{\theta} \sim N_p\left(\boldsymbol{\theta}, \mathrm{diag}\left(\omega_1, \ldots, \omega_p\right)\right)$$

- Accept with probability $\min(1, A)$ where

$$A = \frac{\pi(\boldsymbol{\theta}^*)\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta})} \times \frac{q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)}{q(\boldsymbol{\theta}^*|\boldsymbol{\theta})} = \frac{\pi(\boldsymbol{\theta}^*)\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta})}$$

Sampling $x|\theta, d$

- Blocks of length $m-1$, between observations
- Sample a skeleton path of a conditioned diffusion
- Consider the interval $[t_0, t_1]$

$$(\boldsymbol{x}, \boldsymbol{d}) = \underset{\substack{\uparrow \\ \text{obs}}}{\boldsymbol{x}_0}, \underbrace{\boldsymbol{x}_{\tau_1}, \boldsymbol{x}_{\tau_2}, \ldots, \boldsymbol{x}_{\tau_{m-1}}}_{\text{path (bridge)}}, \underset{\substack{\uparrow \\ \text{obs}}}{\boldsymbol{x}_1}$$

- Distribution of this skeleton path is intractable
- Use a M-H step
- Problem: we need a suitable proposal mechanism

Bridging stategies

- In what follows we drop $\boldsymbol{\theta}$ from notation for simplicity

Pedersen bridge

- Proposed by Pedersen (1995)
- Simulate a path by stepping the Euler-Maruyama approximation forward in time

$$\boldsymbol{X}_{t+\Delta t} = \boldsymbol{X}_t + \Delta \boldsymbol{X}_t, \qquad \Delta \boldsymbol{X}_t \sim N(\boldsymbol{\alpha}(\boldsymbol{X}_t)\Delta t, \boldsymbol{\beta}(\boldsymbol{X}_t)\Delta t)$$

- Target distribution

$$\pi(\boldsymbol{x}_{\tau_1}, \ldots, \boldsymbol{x}_{\tau_{m-1}} | \boldsymbol{x}_{\tau_0}, \boldsymbol{x}_{\tau_m}) \propto \prod_{i=1}^{m} \underbrace{\pi\left(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}}\right)}_{\text{Euler density}}$$

Pedersen bridge

- Acceptance probability of $\min(1, A)$

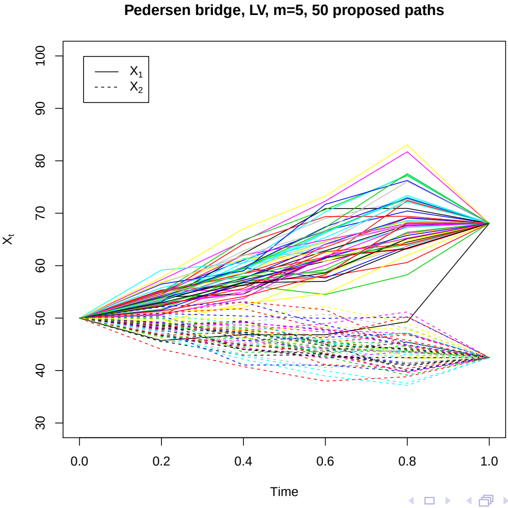
$$A = \underbrace{\frac{\prod_{i=1}^{m} \pi\left(\boldsymbol{x}^*_{\tau_i} | \boldsymbol{x}^*_{\tau_{i-1}}\right)}{\prod_{i=1}^{m} \pi\left(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}}\right)}}_{\text{ratio of target distributions}} \times \underbrace{\frac{\prod_{i=1}^{m-1} \pi\left(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}}\right)}{\prod_{i=1}^{m-1} \pi\left(\boldsymbol{x}^*_{\tau_i} | \boldsymbol{x}^*_{\tau_{i-1}}\right)}}_{\text{ratio of proposal distributions}}$$

$$= \frac{\pi\left(\boldsymbol{x}^*_{\tau_m} | \boldsymbol{x}^*_{\tau_{m-1}}\right)}{\pi\left(\boldsymbol{x}_{\tau_m} | \boldsymbol{x}_{\tau_{m-1}}\right)} \qquad \text{with } \boldsymbol{x}^*_{\tau_m} = \boldsymbol{x}_{\tau_m}$$

  $\boldsymbol{x}^*_\tau$ is the proposed value of the path, $\boldsymbol{x}_\tau$ is the current value of the chain
- For small $\Delta\tau$, A will be close to 0 when $\boldsymbol{x}_{\tau_{m-1}}$ is far from $\boldsymbol{x}_{\tau_m}$
- Problem when $\boldsymbol{x}_{\tau_{m-1}}$ "far" from $\boldsymbol{x}_{\tau_m}$

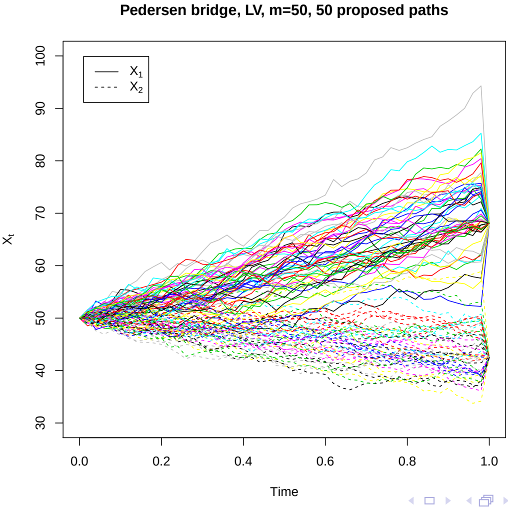
Pedersen bridge

Figure: Pedersen bridge for L-V model $m = 5$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Pedersen bridge

Figure: Pedersen bridge for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Durham & Gallant bridge

- Modified bridge method described by Durham and Gallant (2001)
- Form the proposal density by constructing a Gaussian approximation to $\pi(\boldsymbol{x}_{\tau_i}|\boldsymbol{x}_{\tau_{i-1}}, \boldsymbol{x}_{\tau_m})$
- Using multivariate Normal theory, approximate the joint distribution of $\boldsymbol{X}_{\tau_i}$ and $\boldsymbol{X}_{\tau_m}$ given $\boldsymbol{X}_{\tau_{i-1}} = \boldsymbol{x}_{\tau_{i-1}}$ (and $\boldsymbol{\theta}$) as Gaussian before conditioning on $\boldsymbol{X}_{\tau_m} = \boldsymbol{x}_{\tau_m}$

Durham & Gallant bridge

- The proposal density for $\boldsymbol{X}_{\tau_i}$ is

$$\boldsymbol{X}_{\tau_i}|\boldsymbol{x}_{\tau_{i-1}}, \boldsymbol{x}_{\tau_m} \sim N\left(\boldsymbol{\mu}_{\tau_i}, \boldsymbol{\Sigma}_{\tau_i}\right) \quad i = 1, \ldots, m-1.$$

$$\boldsymbol{\mu}_{\tau_i} = \boldsymbol{x}_{\tau_{i-1}} + \frac{\boldsymbol{x}_{\tau_m} - \boldsymbol{x}_{\tau_{i-1}}}{\tau_m - \tau_{i-1}} \Delta\tau$$

$$\boldsymbol{\Sigma}_{\tau_i} = \frac{\tau_m - \tau_i}{\tau_m - \tau_{i-1}} \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}}) \Delta\tau$$

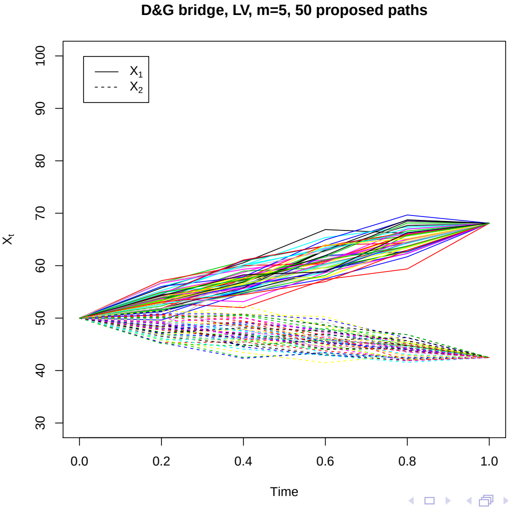
- Acceptance probability is $\min(1, A)$

$$A = \underbrace{\frac{\prod_{i=1}^{m} \pi\left(\boldsymbol{x}^*_{\tau_i} | \boldsymbol{x}^*_{\tau_{i-1}}\right)}{\prod_{i=1}^{m} \pi\left(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}}\right)}}_{\text{ratio of target distributions}} \times \underbrace{\frac{\prod_{i=1}^{m-1} q\left(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}}, \boldsymbol{x}_{\tau_m}\right)}{\prod_{i=1}^{m-1} q\left(\boldsymbol{x}^*_{\tau_i} | \boldsymbol{x}^*_{\tau_{i-1}}, \boldsymbol{x}_{\tau_m}\right)}}_{\text{ratio of proposal distributions}}$$

with $\boldsymbol{x}^*_{\tau_m} = \boldsymbol{x}_{\tau_m}$

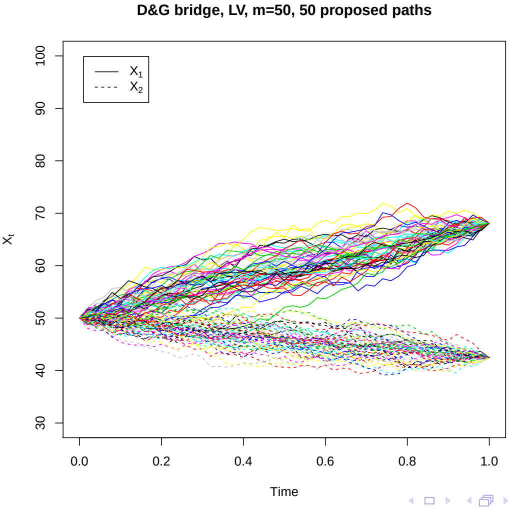
Durham & Gallant bridge

Figure: D&G bridge for L-V model $m = 5$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



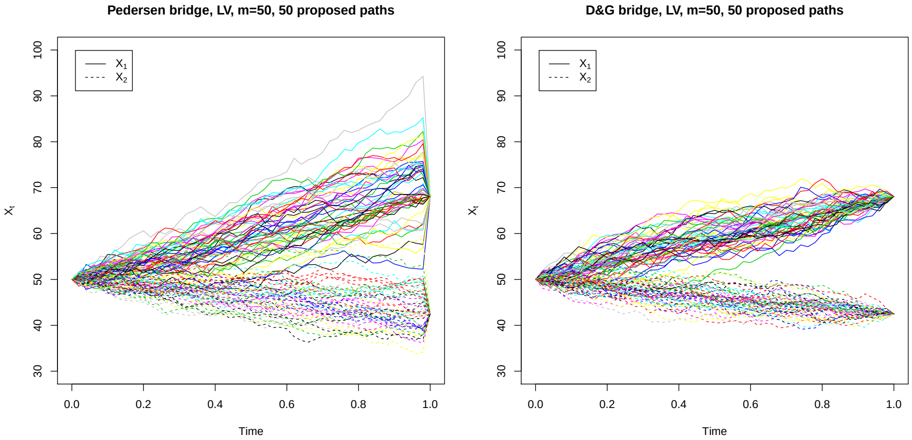
Durham & Gallant bridge

Figure: D&G bridge for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Comparing Pedersen and Durham & Gallant bridges

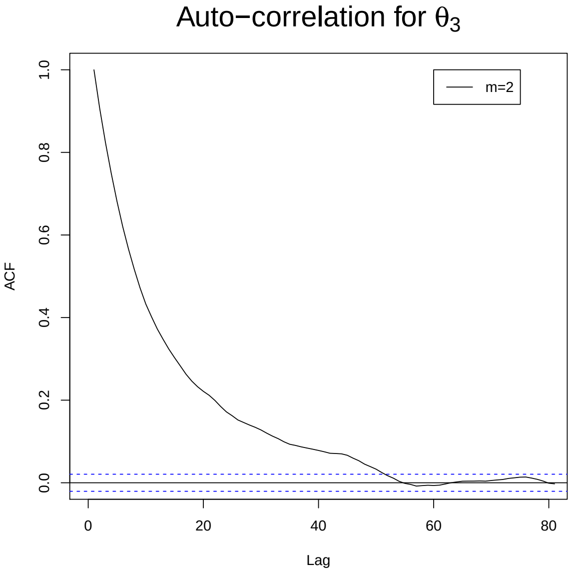
Figure: Pedersen and D&G bridges for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



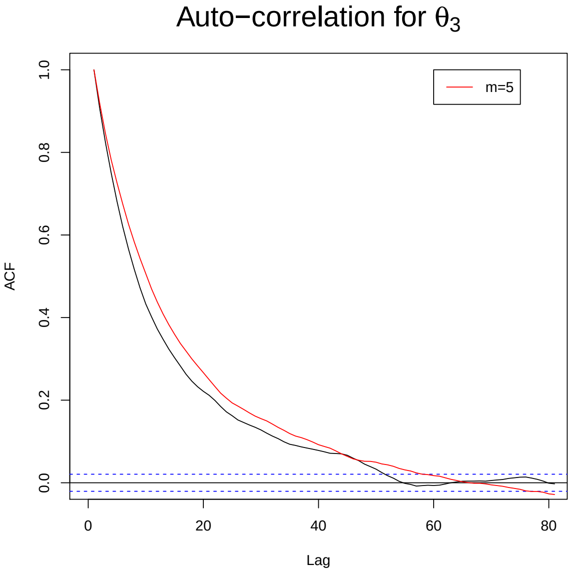
Application: Lotka-Volterra

- Generated synthetic data using the Euler-Maruyama approximation until $T = 50$ with $\boldsymbol{x}_0 = (50, 50)^T$ and time step $\Delta t = 0.01$, thinned to obtain a dataset on a regular grid $0, 1, \ldots, 50$
- Assume prior $\log(\theta_i) \sim U(-7, 2)$ for $i = 1, 2, 3$

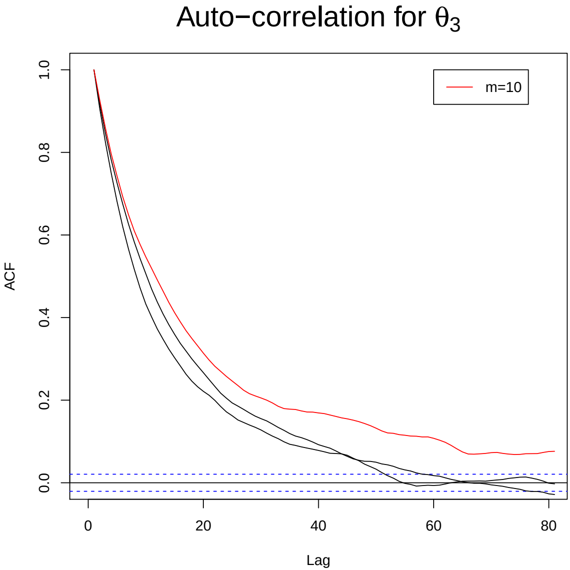
Results - acf plots for L-V model



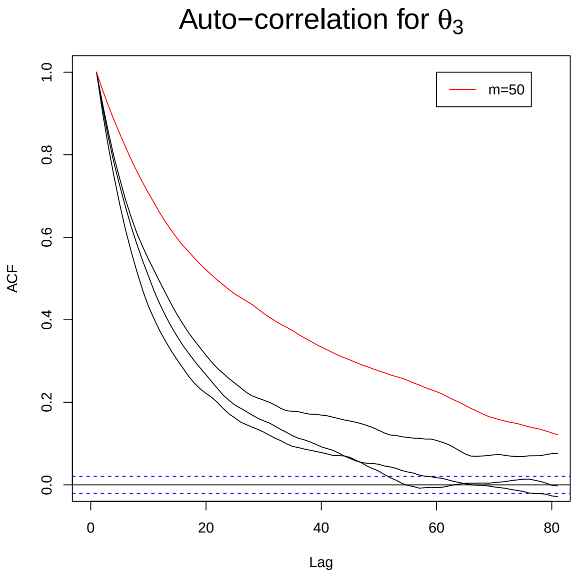
Results - acf plots for L-V model



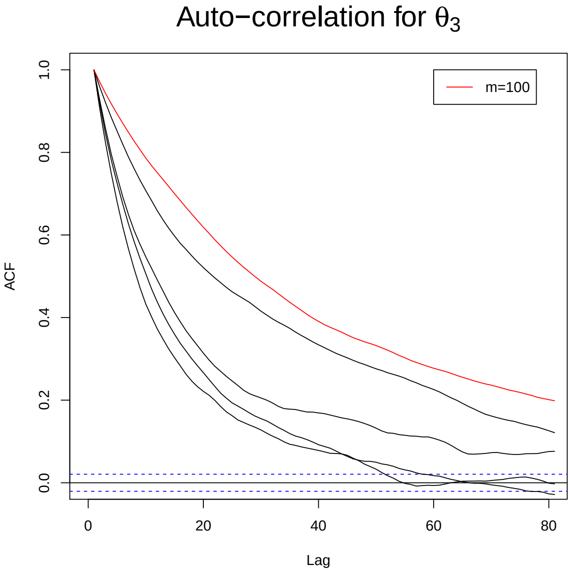
Results - acf plots for L-V model



Results - acf plots for L-V model



Results - acf plots for L-V model



Comments

- If the diffusion coefficient depends on $\boldsymbol{\theta}$, the algorithm is reducible
- For $m \to \infty$, there is an infinite amount of information in the augmented path $(\boldsymbol{x}, \boldsymbol{d})$ about $\boldsymbol{\theta}$
- To see this consider a univariate process satisfying

$$dX_t = \alpha(X_t, \boldsymbol{\theta})\, dt + \sqrt{\boldsymbol{\theta}}\, dW_t$$

For a sample path on times t_0, t_1 the quadratic variation of X_t is

$$\lim_{m\to\infty} \sum_{i=1}^{m} \left[X_{\tau_i} - X_{\tau_{i-1}}\right]^2 = \boldsymbol{\theta}$$

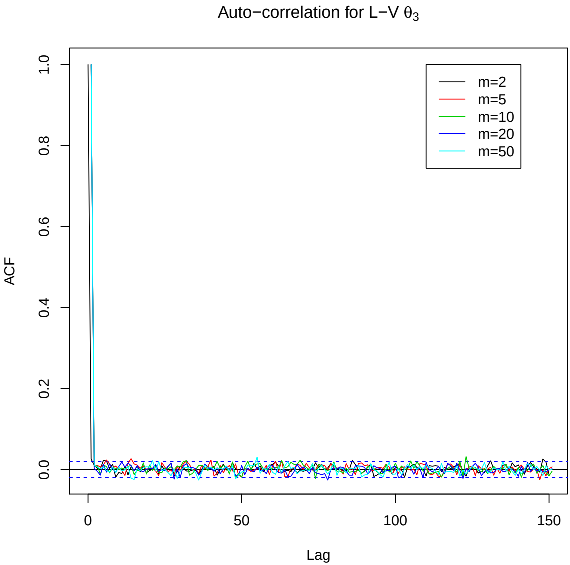
- Naturally, we work with finite discretisations, so the information isn't "infinite", but sufficient to make the algorithm mix very poorly

Future work

- To overcome the problem of poor mixing we will look at using the (Modified) Innovation scheme of Golightly and Wilkinson (2008, 2010)
  - Condition on the Brownian motion innovations that drive, for example the Durham & Gallant construct, to break down the problematic dependence
  - Revised scheme alternates between draws of
    - $\boldsymbol{\theta}|\boldsymbol{w},\boldsymbol{d}$
    - $\boldsymbol{w}|\boldsymbol{\theta},\boldsymbol{d}$

    where $\boldsymbol{w}$ denotes the Brownian increment innovations
- Seek to apply these methods to stochastic differential mixed effects models

Tease



Auto−correlation for L−V θ_3
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
Lag
m=2
m=5
m=10
m=20
m=50

References

- Pedersen, A. R. *A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations.* Scandinavian Journal of Statistics 22 (1) 55-71, 1995
- Durham, G. B. and Gallant, A. R.*Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes.* Journal of Business and Economic Statistics, 20 297-338, 2001
- Roberts, G. O. and Stramer, O. *On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm.* Biometrika, 88 (3) 603-621, 2001
- Golightly, A. Wilkinson, D. J. *Bayesian inference for nonlinear multivariate diffusion models observed with error.* Computational Statistics and Data Analysis, 52 (3) 1674-1693, 2008
- Boys, R. J. Wilkinson, D. J. Kirkwood, T. B. L. *Bayesian inference for a discretely observed stochastic kinetic model.* Statistics and Computing, 18 (2) 125-135, 2008