MCMC schemes for partially observed diffusions

Gavin Whitaker

5th December 2012

Supervisors: RJB and AG

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

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

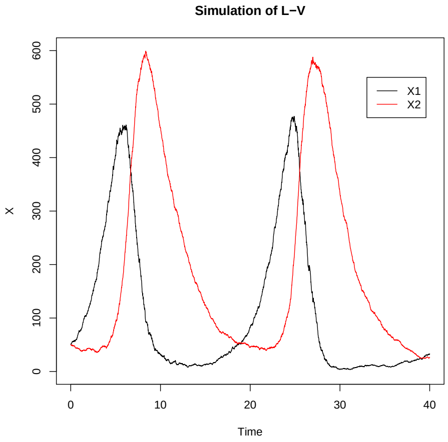
The mass action SDE representation of the system dynamics is given by

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

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

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

Bayesian inference for SDEs

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

- $\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d})$ is proportional to the joint density (1)

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}, \boldsymbol{d}|\boldsymbol{\theta})$$

- 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

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

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

- 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

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 $m = 5$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$

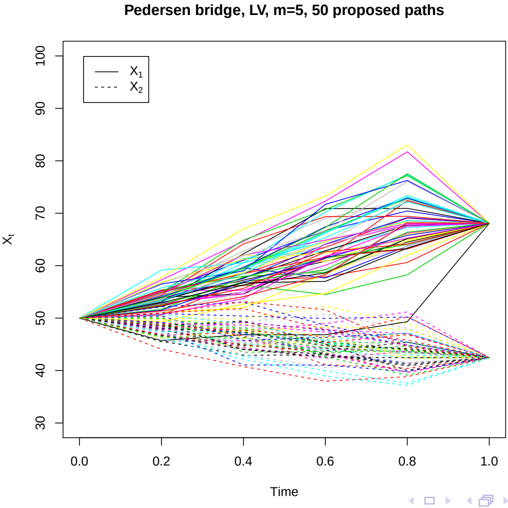
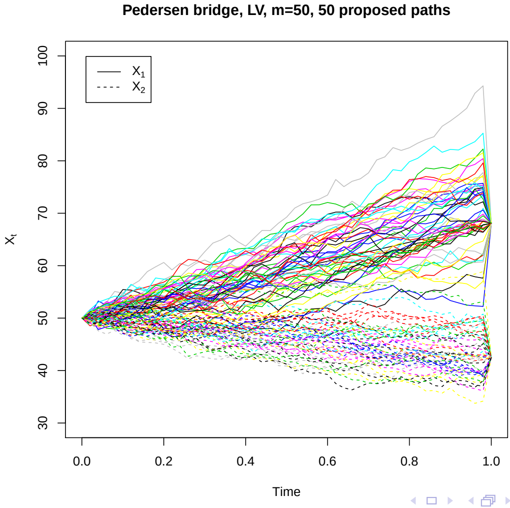


Figure: Pedersen bridge for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Durham & Gallant bridge

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

$$\begin{pmatrix} \boldsymbol{X}_{\tau_i} \\ \boldsymbol{X}_{\tau_m} \end{pmatrix} \Bigg| \boldsymbol{x}_{\tau_{i-1}} \sim N_{2d}\left(\boldsymbol{\mu}^*_{\tau_i}, \boldsymbol{\Sigma}^*_{\tau_i}\right)$$

$$\boldsymbol{\mu}^*_{\tau_i} = \begin{pmatrix} \boldsymbol{x}_{\tau_{i-1}} + \boldsymbol{\alpha}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau \\ \boldsymbol{x}_{\tau_{i-1}} + \boldsymbol{\alpha}(\boldsymbol{x}_{\tau_{i-1}})\Delta^- \end{pmatrix}$$

$$\boldsymbol{\Sigma}^*_{\tau_i} = \begin{pmatrix} \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau & \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau \\ \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau & \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})\Delta^- \end{pmatrix}$$

where $\Delta^- = \tau_m - \tau_{i-1}$

- We use two levels of approximation
  1. $\boldsymbol{X}_{\tau_m} | \boldsymbol{x}_{\tau_i} \sim N(\boldsymbol{x}_{\tau_i} + \boldsymbol{\alpha}(\boldsymbol{x}_{\tau_i})(\tau_m - \tau_i), \boldsymbol{\beta}(\boldsymbol{x}_{\tau_i})(\tau_m - \tau_i))$
  2. $\boldsymbol{\alpha}(\boldsymbol{x}_{\tau_i})$ and $\boldsymbol{\beta}(\boldsymbol{x}_{\tau_i})$ are replaced by $\boldsymbol{\alpha}(\boldsymbol{x}_{\tau_{i-1}})$ and $\boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})$ to obtain a linear Gaussian structure

Durham & Gallant bridge

- Now condition on $\boldsymbol{X}_{\tau_m} = \boldsymbol{x}_{\tau_m}$, giving the proposal density for $\boldsymbol{X}_{\tau_i}$ as

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

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

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

Durham & Gallant bridge

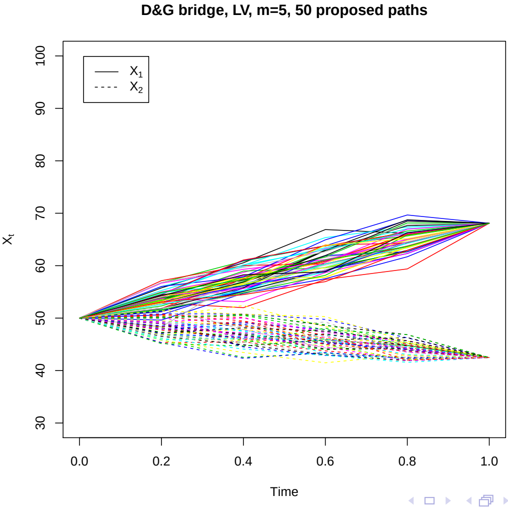
- Acceptance probability is min$\{1, A\}$

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

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

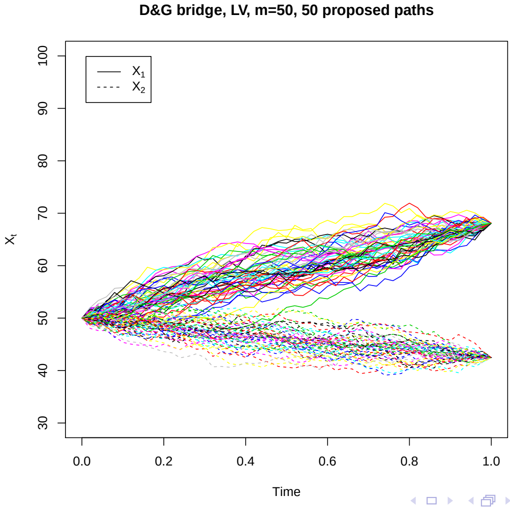
Durham & Gallant bridge

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



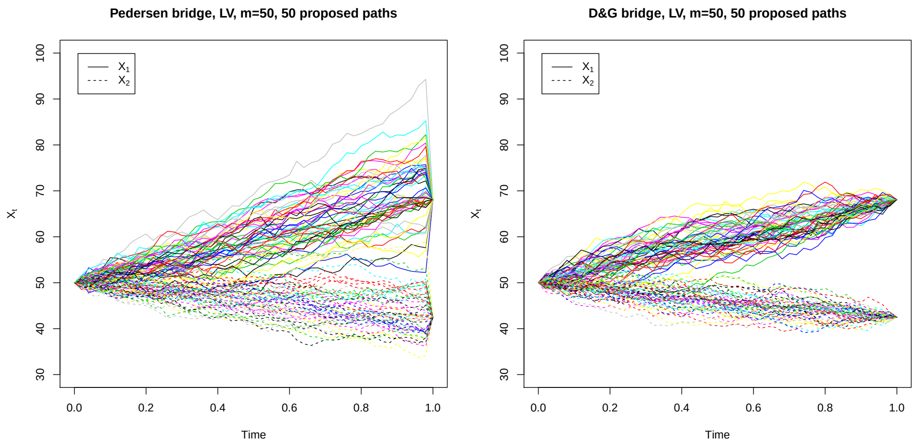
Durham & Gallant bridge

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 $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Comparing Pedersen and Durham & Gallant bridges

Table: Empirical acceptance probabilities based on 50 000 simulations for the Pedersen and the Durham and Gallant bridges, L-V model, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$

Scheme	Empirical acceptance probability					
	$m = 2$	$m = 5$	$m = 10$	$m = 20$	$m = 50$	$m = 100$
Pedersen	0.610	0.262	0.127	0.057	0.018	0.008
D & G	0.764	0.724	0.717	0.722	0.718	0.716

Modified diffusion bridge for partial observation

Partial observation

- Partition $\boldsymbol{X}_t$ as $\boldsymbol{X}_t = \begin{pmatrix} \boldsymbol{X}_t^o \\ \boldsymbol{X}_t^u \end{pmatrix}$
- Component $\boldsymbol{X}_t^o$ observed at times $t_0, t_1, \ldots, t_n$
- Component $\boldsymbol{X}_t^u$ not observed anywhere
- Sample $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}$ using overlapping blocks, over intervals $[t_{i-1}, t_{i+1}]$ $i = 1, \ldots, n-1$

Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved

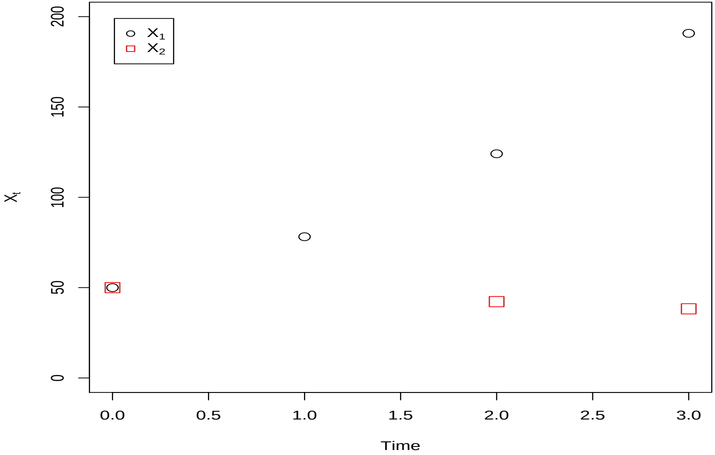
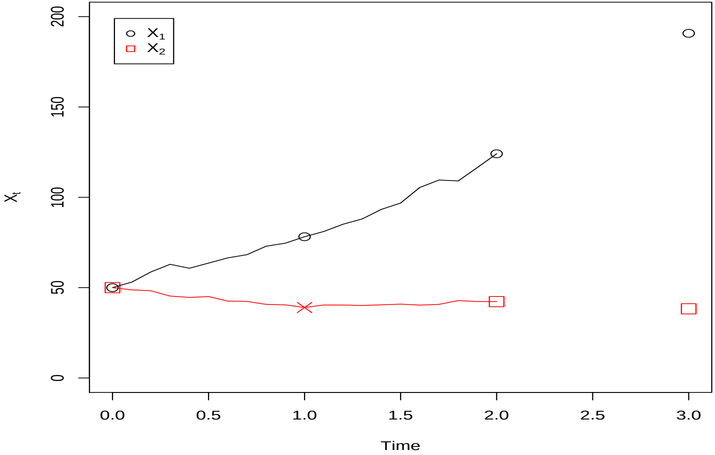
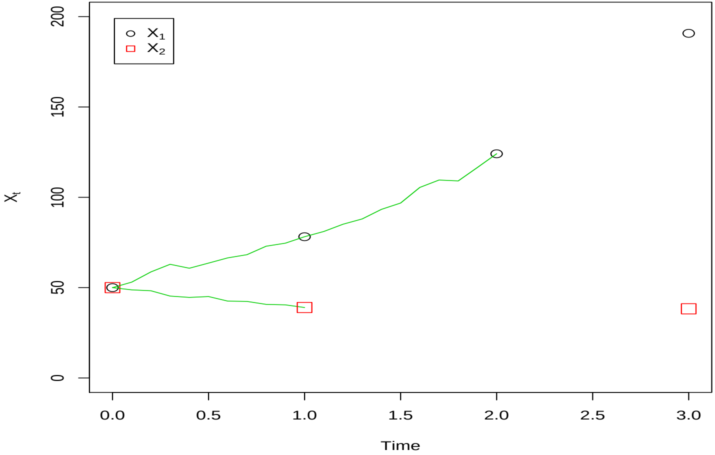


Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved



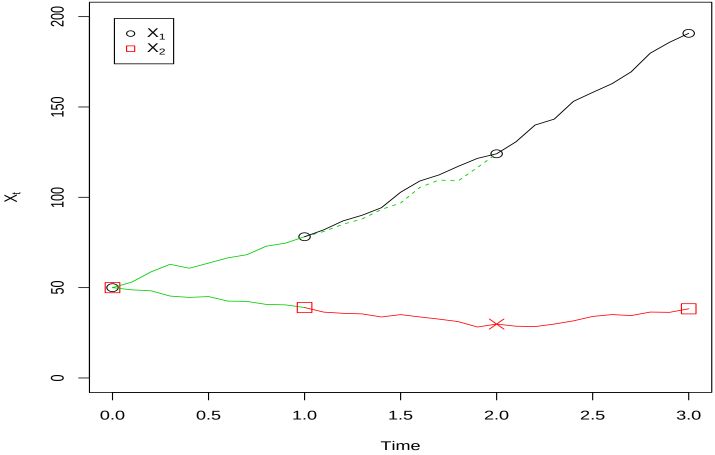
Modified diffusion bridge for partial observation

Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved



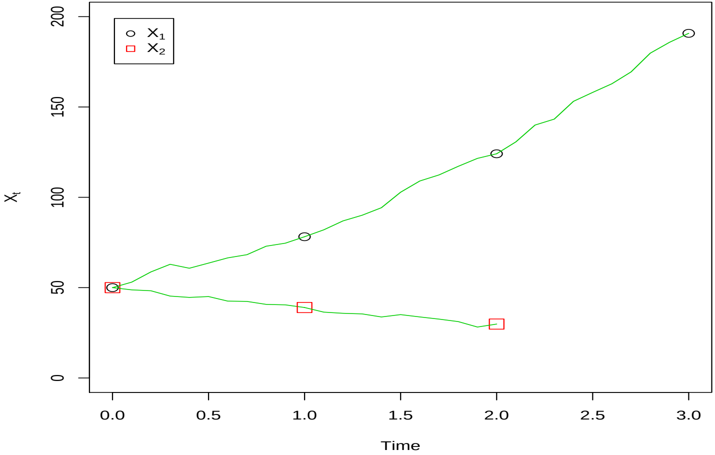
Modified diffusion bridge for partial observation

Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved



Modified diffusion bridge for partial observation

Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved



Modified diffusion bridge for partial observation

Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved

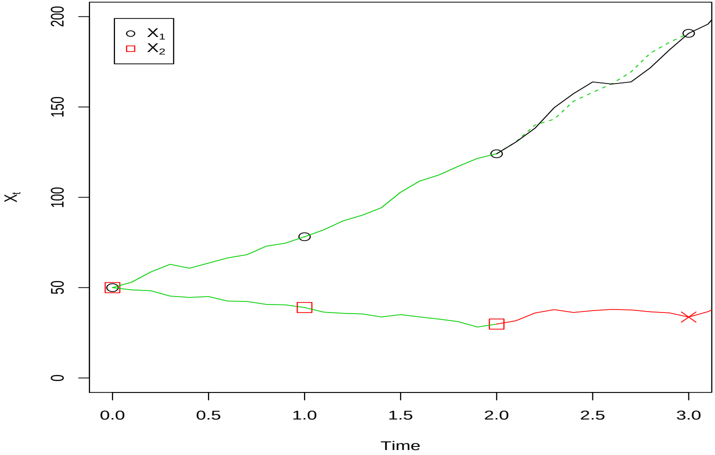
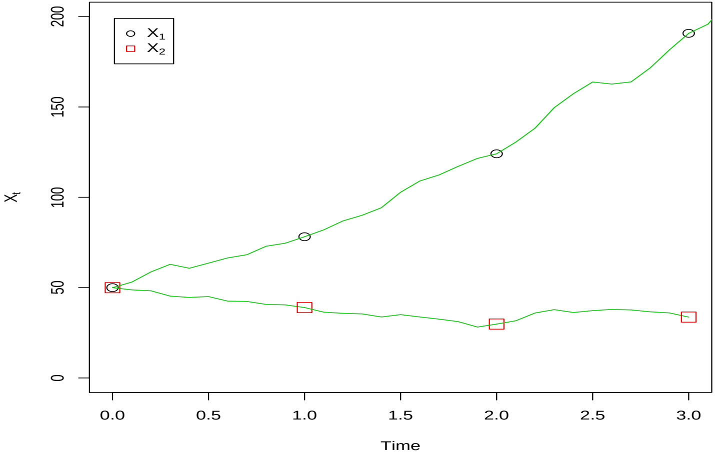


Figure: Sample path for partial observations, $m = 10$. X_1 observed, X_2 unobserved



Partial observation

- Consider an interval $[t_0, t_2]$ and recall the partition

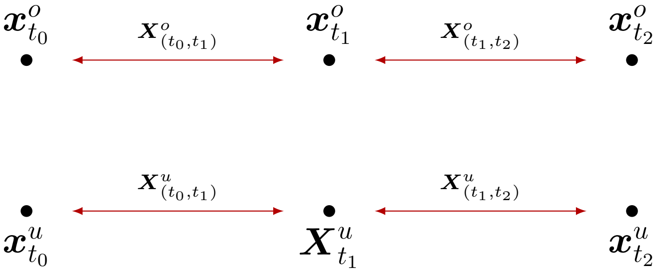
$$t_0 = \tau_0 < \tau_1 < \ldots < \tau_m = t_1 < \tau_{m+1} < \ldots < \tau_{2m} = t_2$$

- Target distribution given by

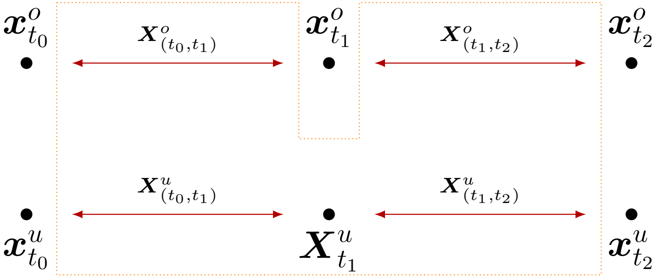
$$\pi(\boldsymbol{x}_{(t_0,t_1)}, \boldsymbol{x}^u_{t_1}, \boldsymbol{x}_{(t_1,t_2)} | \boldsymbol{x}_{\tau_0}, \boldsymbol{x}_{\tau_{2m}}, \boldsymbol{x}^o_{\tau_m}) \propto \prod_{i=1}^{2m} \pi(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}})$$

where $\boldsymbol{x}_{(t_0,t_1)} = (\boldsymbol{x}_{\tau_1}, \ldots, \boldsymbol{x}_{\tau_{m-1}})$

Partial observation

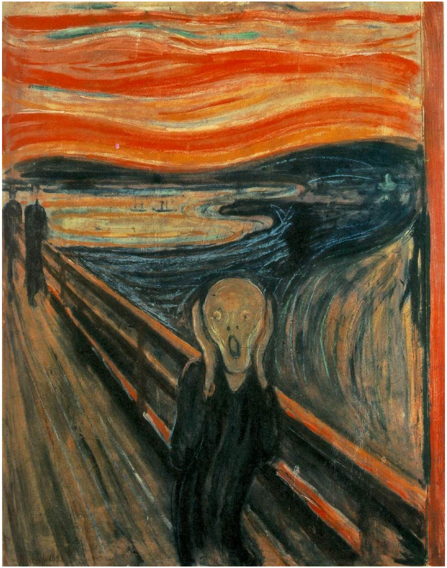


Possible proposals



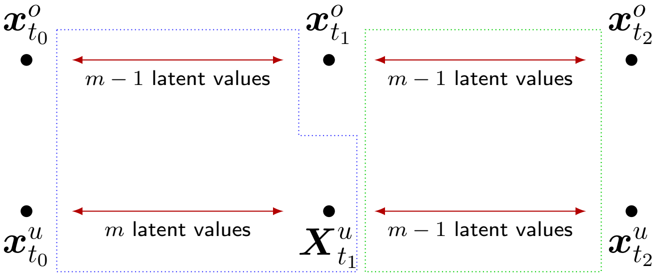
Idealised case

Partial observation

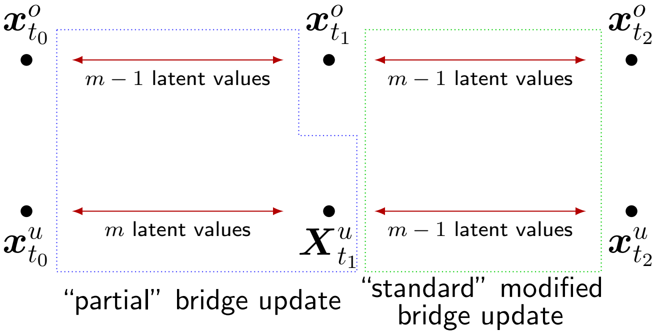


Partial observation

Possible proposals



Possible proposals



Modified diffusion bridge for partial observation

Approximate the joint distribution of $\boldsymbol{X}^o_{\tau_i}$, $\boldsymbol{X}^u_{\tau_i}$ and $\boldsymbol{X}^o_{\tau_m}$ given $\boldsymbol{X}_{\tau_{i-1}} = \boldsymbol{x}_{\tau_{i-1}}$ (and $\boldsymbol{\theta}$) as Gaussian and condition on $\boldsymbol{X}^o_{\tau_m} = \boldsymbol{x}^o_{\tau_m}$

$$\begin{pmatrix} \boldsymbol{X}^o_{\tau_i} \\ \boldsymbol{X}^u_{\tau_i} \end{pmatrix} \Bigg| \boldsymbol{x}_{\tau_{i-1}}, \boldsymbol{x}^o_{\tau_m} \sim N_d\left(\boldsymbol{\mu}_{\tau_{i-1}}, \boldsymbol{\Sigma}_{\tau_{i-1}}\right)$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\tau_{i-1}} = \boldsymbol{x}_{\tau_{i-1}} + \boldsymbol{\alpha}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau + \begin{pmatrix} \boldsymbol{\beta}^o(\boldsymbol{x}_{\tau_{i-1}}) \\ \boldsymbol{\beta}^{uo}(\boldsymbol{x}_{\tau_{i-1}}) \end{pmatrix} \Delta\tau \left(\boldsymbol{\beta}^o(\boldsymbol{x}_{\tau_{i-1}})\Delta^-\right)^{-1} \\ \times \left\{\boldsymbol{x}^o_{\tau_m} - \left(\boldsymbol{x}^o_{\tau_{i-1}} + \boldsymbol{\alpha}^o(\boldsymbol{x}_{\tau_{i-1}})\Delta^-\right)\right\}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\Sigma}_{\tau_{i-1}} = \boldsymbol{\beta}(\boldsymbol{x}_{\tau_{i-1}})\Delta\tau - \begin{pmatrix} \boldsymbol{\beta}^o(\boldsymbol{x}_{\tau_{i-1}}) \\ \boldsymbol{\beta}^{uo}(\boldsymbol{x}_{\tau_{i-1}}) \end{pmatrix} \Delta\tau \left(\boldsymbol{\beta}^o(\boldsymbol{x}_{\tau_{i-1}})\Delta^-\right)^{-1} \\ \times \left(\boldsymbol{\beta}^o(\boldsymbol{x}_{\tau_{i-1}}), \boldsymbol{\beta}^{ou}(\boldsymbol{x}_{\tau_{i-1}})\right)\Delta\tau\end{aligned}$$

Modified diffusion bridge for partial observation

- Target distribution

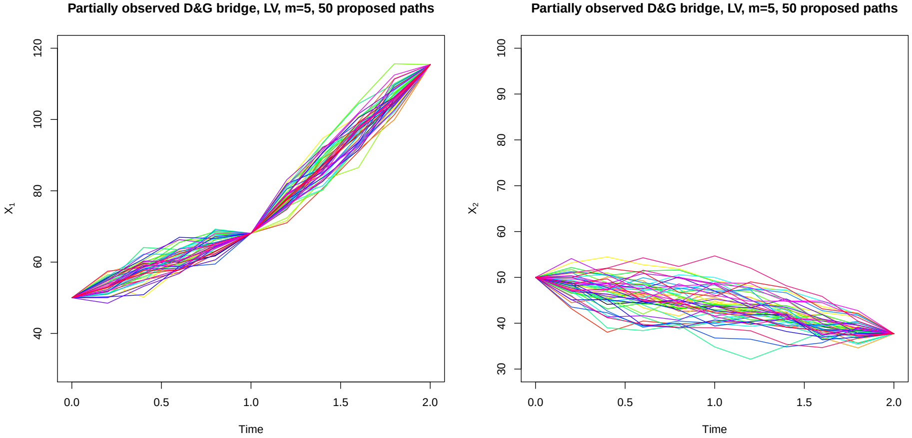
$$\pi(\boldsymbol{x}_{(t_0,t_1)}, \boldsymbol{x}^u_{t_1}, \boldsymbol{x}_{(t_1,t_2)} | \boldsymbol{x}_{\tau_0}, \boldsymbol{x}_{\tau_{2m}}, \boldsymbol{x}^o_{\tau_m}) \propto \prod_{i=1}^{2m} \pi(\boldsymbol{x}_{\tau_i} | \boldsymbol{x}_{\tau_{i-1}})$$

- Proposal mechanism:

1. Propose $\boldsymbol{x}_{(t_0,t_1)}$ using the modified bridge for partial observations
2. Propose $\boldsymbol{x}^u_{\tau_m}$ from $q(\boldsymbol{x}^u_{\tau_m} | \boldsymbol{x}_{\tau_{m-1}}, \boldsymbol{x}^o_{\tau_m})$
3. Propose $\boldsymbol{x}_{(t_1,t_2)}$ using the "standard" modified bridge

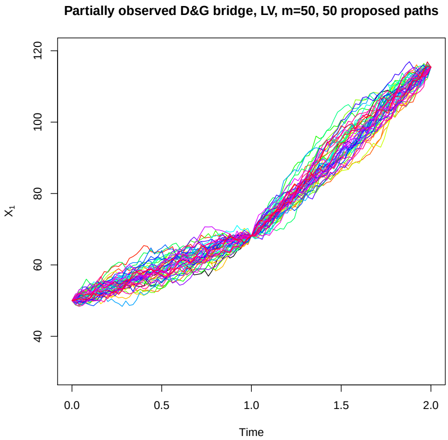
Modified diffusion bridge for partial observation

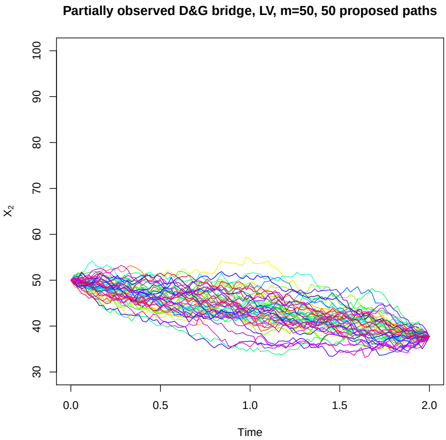
Figure: Partially observed D&G bridges for L-V model $m = 5$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1^o = 68.09$, $\boldsymbol{x}_2 = (115.41, 37.718)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



Modified diffusion bridge for partial observation

Figure: Partially observed D&G bridges for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1^o = 68.09$, $\boldsymbol{x}_2 = (115.41, 37.718)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$

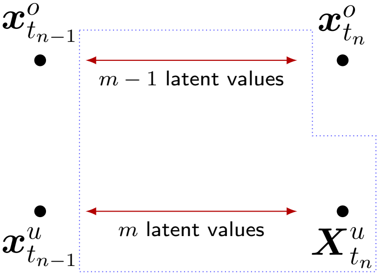




Modified diffusion bridge for partial observation

Update the end

- Consider an interval of length m, over $[t_{n-1}, t_n]$

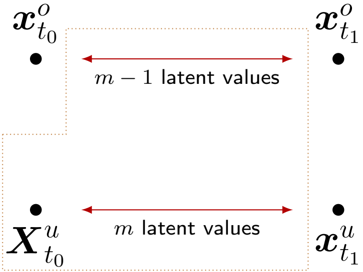


- $\boldsymbol{X}_{t_{n-1}} = \boldsymbol{x}_{t_{n-1}}$ and $\boldsymbol{X}^o_{t_n} = \boldsymbol{x}^o_{t_n}$ fixed
- Use M-H step with proposal $q\left(\boldsymbol{x}_{(t_{n-1},t_n)}, \boldsymbol{x}^u_{t_n} | \boldsymbol{x}_{t_{n-1}}, \boldsymbol{x}^o_{t_n}\right)$

Modified diffusion bridge for partial observation

Update the start

- ▷ Consider an interval of length m, over $[t_0, t_1]$



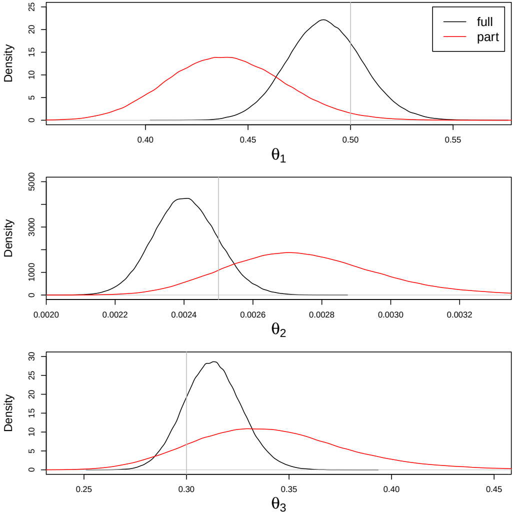
- ▷ $\boldsymbol{X}_{t_1} = \boldsymbol{x}_{t_1}$ and $\boldsymbol{X}^o_{t_0} = \boldsymbol{x}^o_{t_0}$ fixed
- ▷ Using M-H step, draw $\boldsymbol{x}^u_{t_0}$ from $\pi(\boldsymbol{x}^u_{t_0})$ and update the path with proposal $q\left(\boldsymbol{x}_{(t_0,t_1)}, |\boldsymbol{x}_{t_0}, \boldsymbol{x}_{t_1}\right)$
- ▷ Could also perform a random walk on $\boldsymbol{x}^u_{t_0}$

Application: Lotka-Volterra

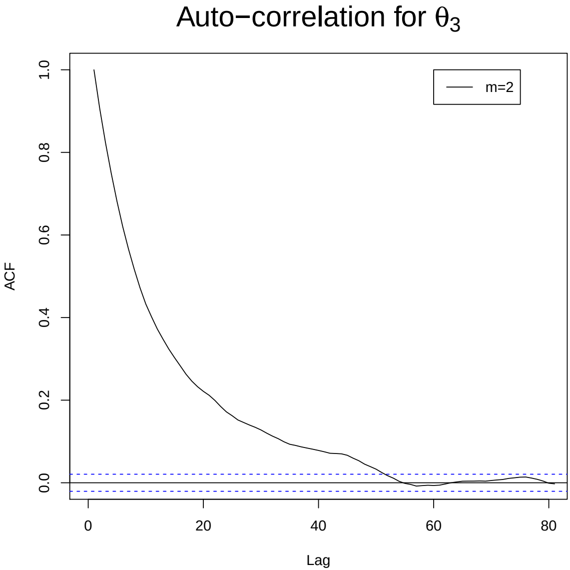
- Generated synthetic data using the Euler-Maruyama approximation until $T = 50$ with $\boldsymbol{x}_0 = (50, 50)^T$ and time step $\Delta t = 0.01$, thinned to obtain a dataset on a regular grid $0, 1, \ldots, 50$
- Assume we do not observe $\boldsymbol{X}_2$ for the partially observed case
- Assume prior $\log(\boldsymbol{\theta}_i) \sim U(-7, 2)$ for $i = 1, 2, 3$ and $\boldsymbol{x}^u_{t_0} \sim U(45, 55)$

Results - kernel density plots for L-V model

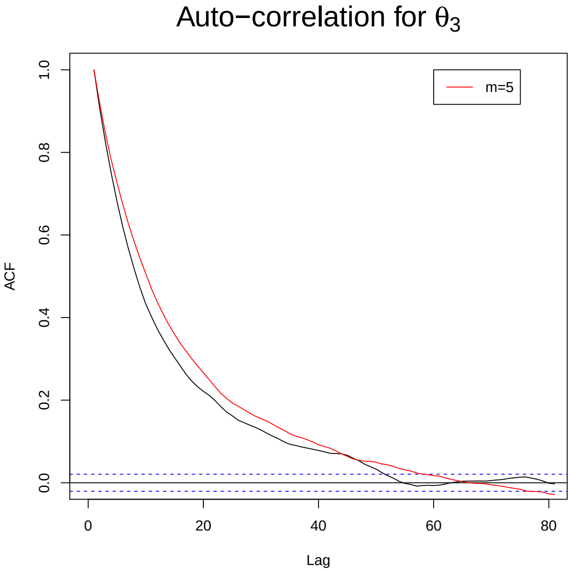
Figure: $m = 5$, 1 million iterations



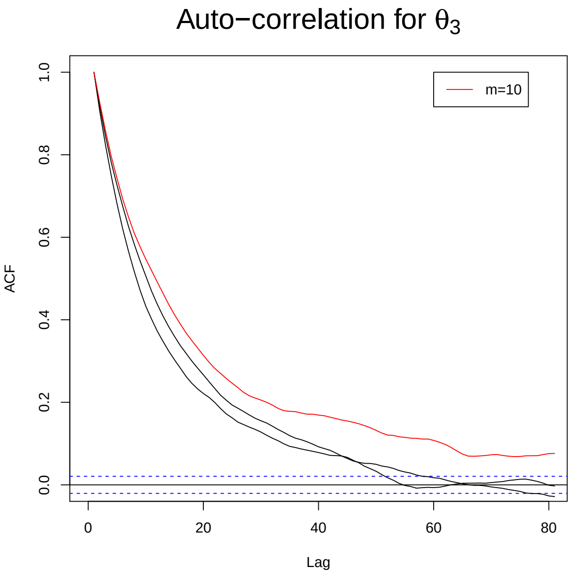
Results - acf plots for L-V model



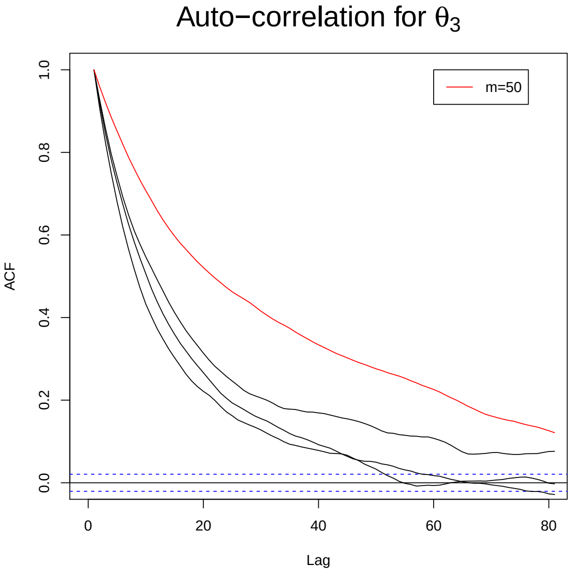
Results - acf plots for L-V model



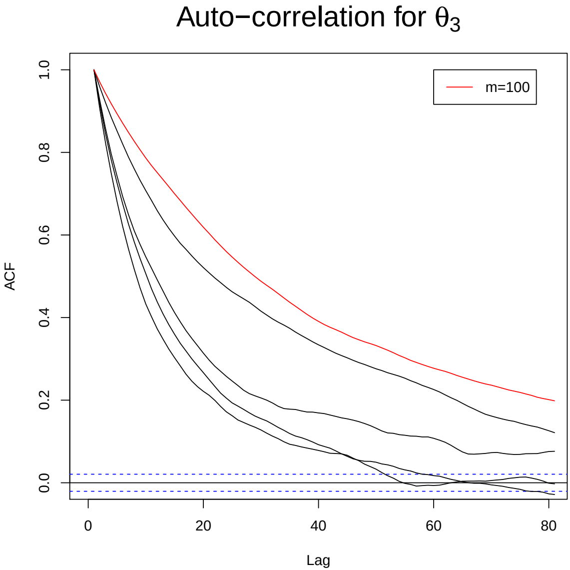
Results - acf plots for L-V model



Results - acf plots for L-V model



Results - acf plots for L-V model



Results - empirical acceptance probabilities

Table: Empirical acceptance probabilities for the L-V model, (fully observed case)

Empirical acceptance probability				
$m=2$	$m=5$	$m=10$	$m=50$	$m=100$
0.2871	0.2768	0.2651	0.2030	0.1544

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

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

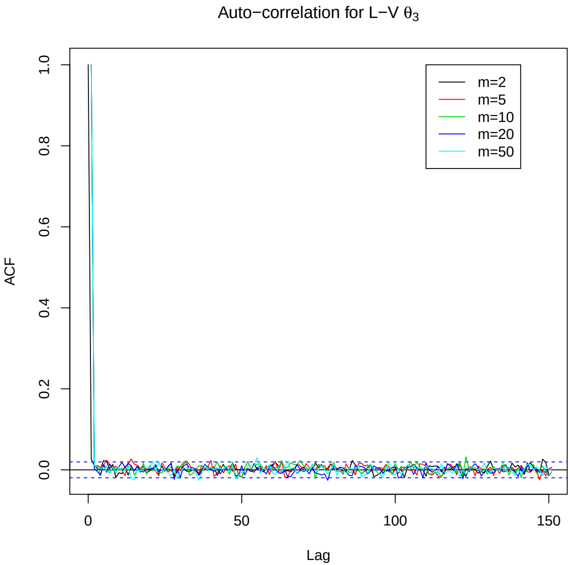
- Naturally, we work with finite discretisations, so the information isn't "infinite", but sufficient to make the algorithm mix very poorly

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.
  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, ~~2000~~, ~~2004~~, 2008