An innovative look at MCMC schemes for partially observed diffusions

Gavin Whitaker

13th March 2013

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

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

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

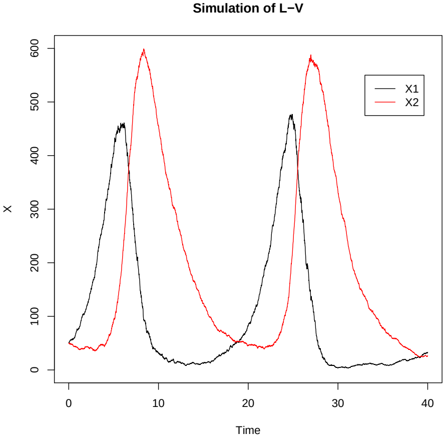


Figure: Numerical solution for L-V model, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$

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

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

Bayesian inference for SDEs

Formulate joint posterior for parameters and latent data as

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

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

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

- 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

Durham & Gallant bridge

- Modified bridge method described by Durham and Gallant (2001)

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

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

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

- See my SBSSB talk on 5th December 2012

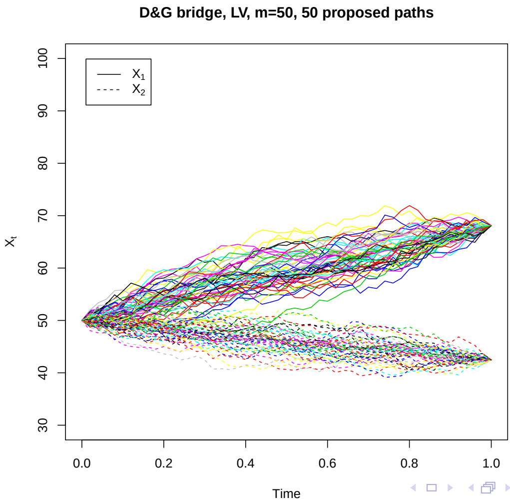
Durham & Gallant bridge

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

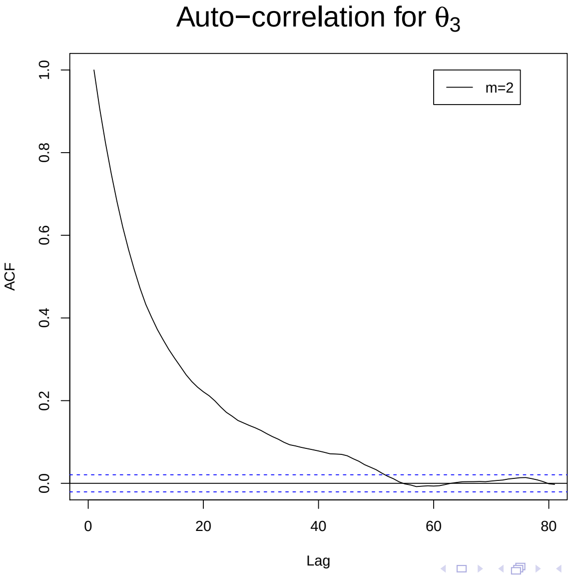
Figure: D&G bridge for L-V model $m = 50$, $\boldsymbol{x}_0 = (50, 50)^T$, $\boldsymbol{x}_1 = (68.09, 42.48)^T$, $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)^T$



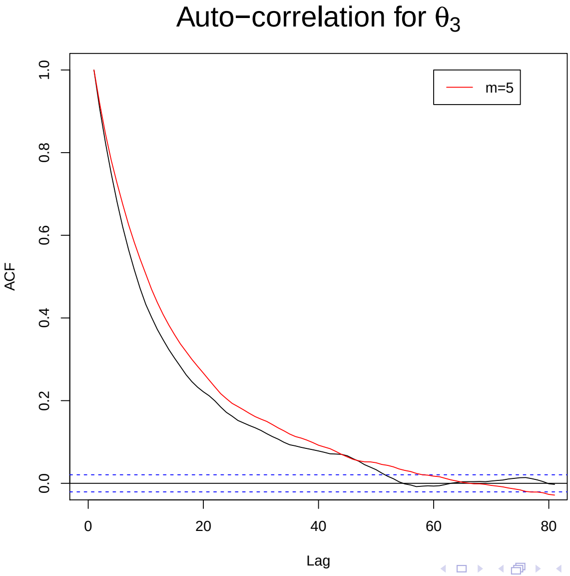
Application: Lotka-Volterra

- 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 $i = 1, 2, 3$

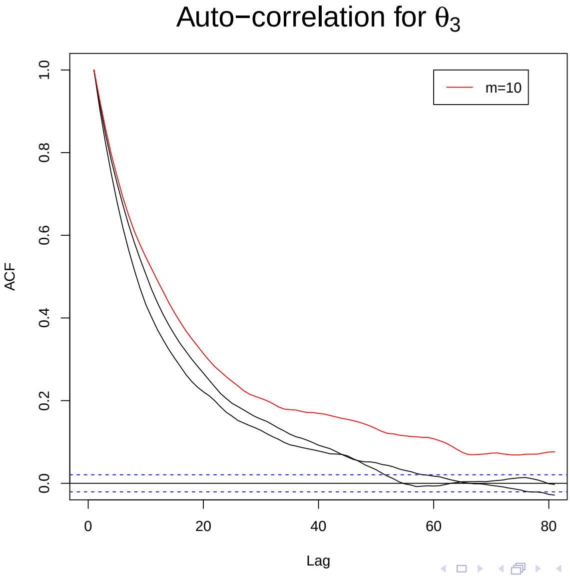
Acf plots for L-V model



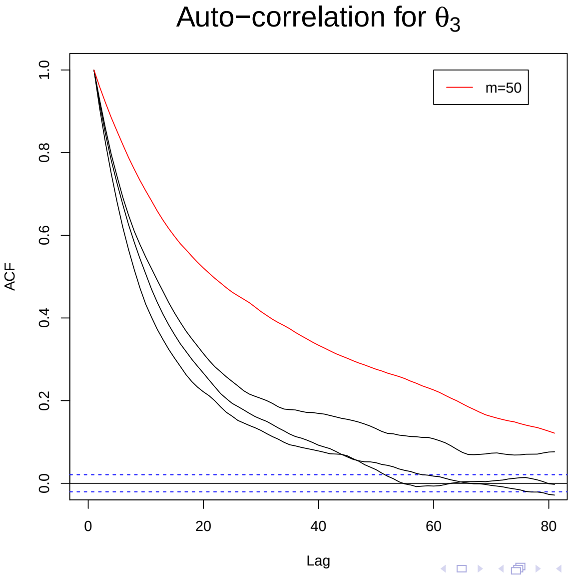
Acf plots for L-V model



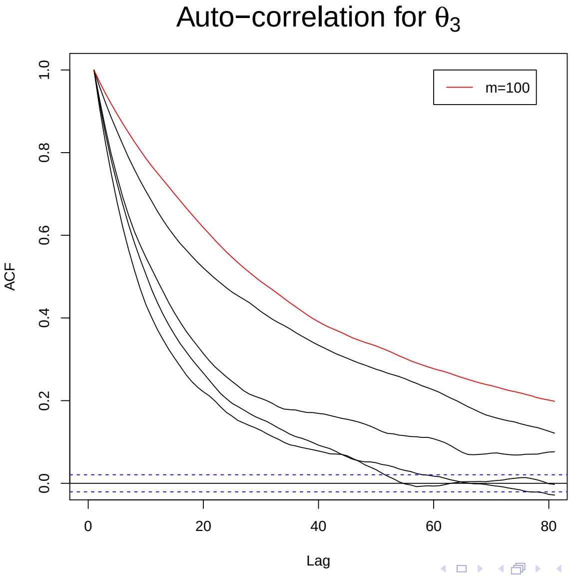
Acf plots for L-V model



Acf plots for L-V model



Acf plots for L-V model



Addressing the mixing

- Use the (Modified) Innovation scheme of Golightly and Wilkinson (2008, 2010)
- Recall that when we use a Gibbs Sampling approach we alternate between draws of

  1. $\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{d}$
  2. $\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{d}$

  where in 1, we sample from the target distribution

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

(Modified) Innovation scheme

- Under the Durham and Gallant bridge

$$\Delta \boldsymbol{X}_{\tau_i} = \boldsymbol{\alpha}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta})\Delta\tau + \boldsymbol{\beta}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta})^{\frac{1}{2}}\Delta \boldsymbol{W}_{\tau_i}$$

- Therefore

$$\Delta \boldsymbol{W}_{\tau_i} = \boldsymbol{\beta}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta})^{-\frac{1}{2}}\left\{\Delta \boldsymbol{X}_{\tau_i} - \boldsymbol{\alpha}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta})\Delta\tau\right\}$$

where

$$\boldsymbol{\alpha}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta}) = \frac{\boldsymbol{X}_{\tau_m} - \boldsymbol{X}_{\tau_{i-1}}}{\tau_m - \tau_{i-1}}$$

and

$$\boldsymbol{\beta}^*(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta}) = \frac{\tau_m - \tau_i}{\tau_m - \tau_{i-1}}\boldsymbol{\beta}(\boldsymbol{X}_{\tau_{i-1}}, \boldsymbol{\theta})$$

(Modified) Innovation scheme

- Let

$$\boldsymbol{w} = (\Delta \boldsymbol{w}_{\tau_1}, \Delta \boldsymbol{w}_{\tau_2}, \ldots, \Delta \boldsymbol{w}_{\tau_{m-1}}, \Delta \boldsymbol{w}_{\tau_{m+1}}, \ldots, \ldots, \Delta \boldsymbol{w}_{\tau_{nm-1}})$$

  denote the Brownian increment innovations
- We now alternate between draws of
  - $\boldsymbol{\theta} | \boldsymbol{w}, \boldsymbol{d}$
  - $\boldsymbol{w} | \boldsymbol{\theta}, \boldsymbol{d}$

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

- The target distribution is

$$\pi(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{d}) \propto \pi(\boldsymbol{\theta})\, \pi\{g(\boldsymbol{w}, \boldsymbol{\theta})|\boldsymbol{d}\}\, J(\boldsymbol{\theta})$$

where the Jacobian for one increment is

$$J(\boldsymbol{\theta}) = \left| \frac{\partial \Delta \boldsymbol{W}_t}{\partial \Delta \boldsymbol{X}_t} \right| = |\boldsymbol{\beta}^*(\boldsymbol{X}_t, \boldsymbol{\theta})|^{-\frac{1}{2}}$$

- The target distribution therefore becomes

$$\pi(\boldsymbol{\theta}|\boldsymbol{w}, \boldsymbol{d}) \propto \pi(\boldsymbol{\theta}) \prod_{i=0}^{nm-1} \pi(\boldsymbol{X}_{\tau_{i+1}}|\boldsymbol{X}_{\tau_i}, \boldsymbol{\theta}) \times \prod_{i=0}^{n-1} \prod_{j=0}^{m-2} |\boldsymbol{\beta}^*(\boldsymbol{X}_{t_i+\tau_j}, \boldsymbol{\theta})|^{-\frac{1}{2}}$$

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

- Typically intractable so use a M-H step to get $\boldsymbol{\theta}^*$
- We construct a new path $\boldsymbol{x}^* = g(\boldsymbol{w}, \boldsymbol{\theta}^*)$ by applying the Durham and Gallant bridge
- Accept with probability

$$A = \frac{\pi(\boldsymbol{\theta}^*)\, \pi\{g(\boldsymbol{w}, \boldsymbol{\theta}^*)|\boldsymbol{d}\}\, J(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})\, \pi\{g(\boldsymbol{w}, \boldsymbol{\theta})|\boldsymbol{d}\}\, J(\boldsymbol{\theta})}$$

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

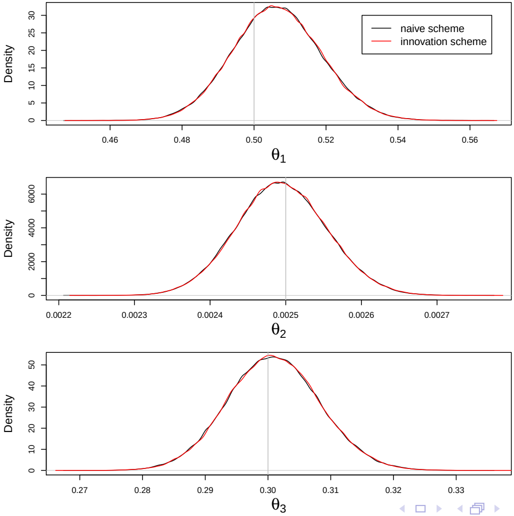
- Use a M-H step with the Durham and Gallant bridge as its proposal method
- Using the one to one relationship between $\boldsymbol{w}$ and $\boldsymbol{x}$ under the Durham and Gallant bridge, we can sample $\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{d}$ to give us realisations from $\boldsymbol{w}|\boldsymbol{\theta},\boldsymbol{d}$
- This is just the same step as used in the "naive" scheme
- See my SBSSB talk on 5th December 2012

Application: Lotka-Volterra

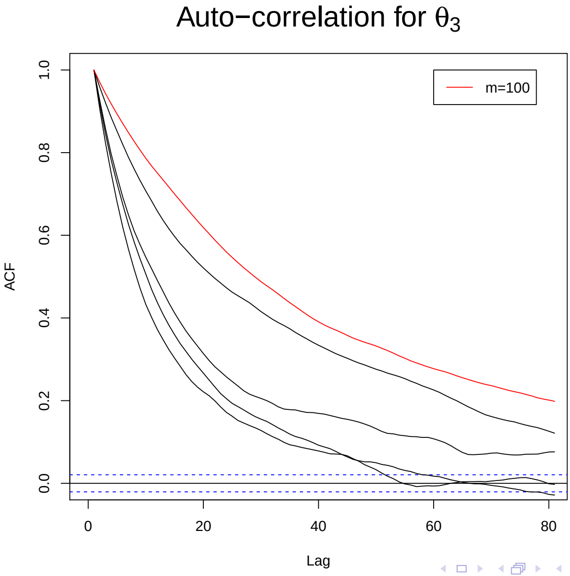
- Generated synthetic data using the Euler-Maruyama approximation until $T = 50$ with $\boldsymbol{x}_0 = (50, 50)^T$ and time step $\Delta t = 0.01$, thinned to obtain a dataset on a regular grid $0, 1, \ldots, 50$
- Assume prior $\log \theta_i \sim U(-7, 2)$ for $i = 1, 2, 3$

Kernel density plots for L-V model

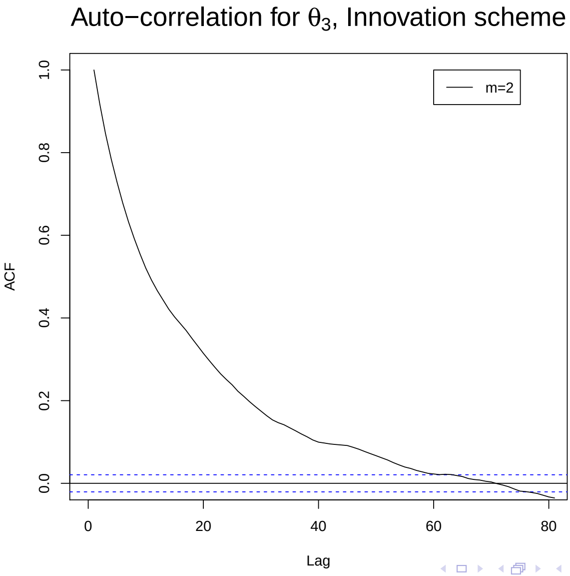
Figure: $m = 5$, 1 million iterations



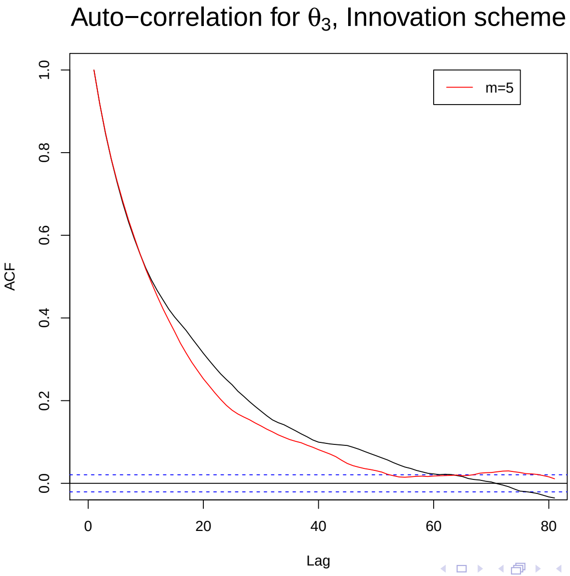
Acf plots for L-V model - Reminder



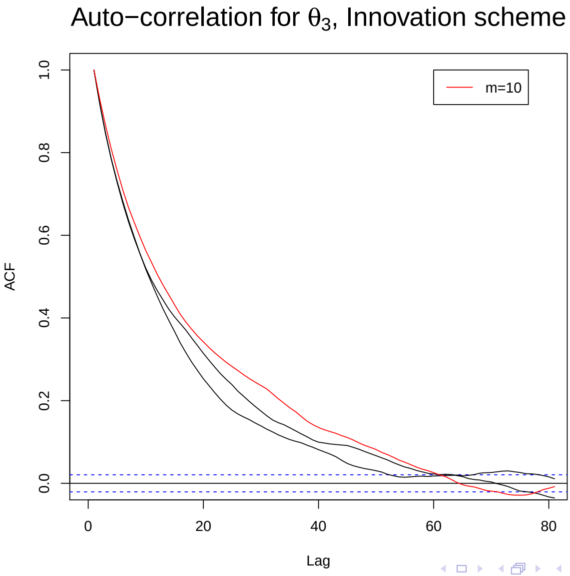
Acf plots for L-V model, Innovation scheme



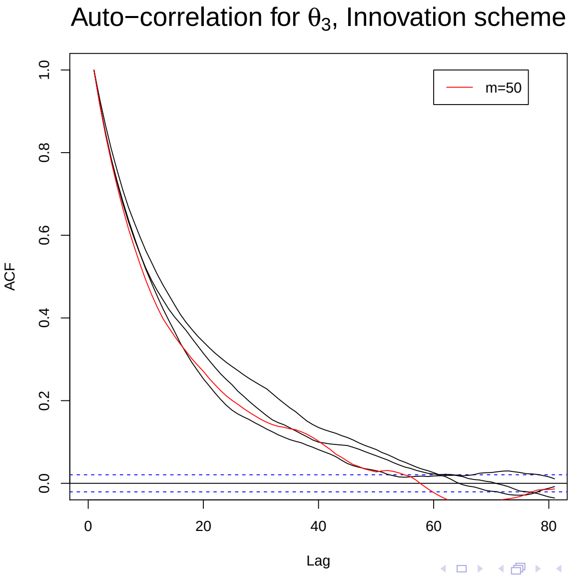
Acf plots for L-V model, Innovation scheme



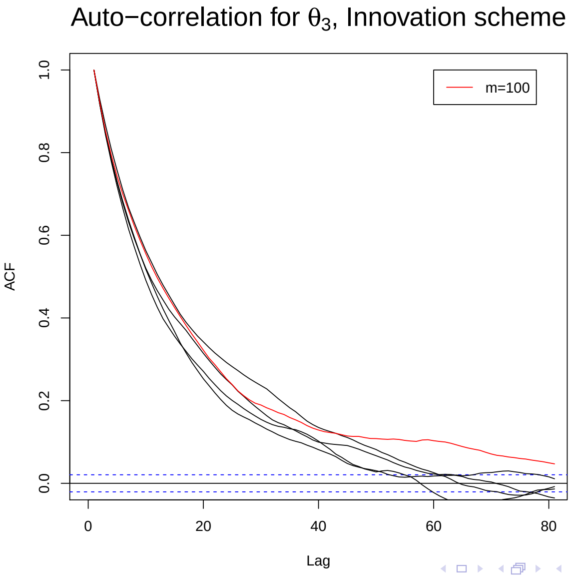
Acf plots for L-V model, Innovation scheme



Acf plots for L-V model, Innovation scheme



Acf plots for L-V model, Innovation scheme



Why the improvement?

- Under the "naive" scheme as $m \to \infty$

$$\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta}) \neq \boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta}^*) \Rightarrow L(\boldsymbol{\theta}, \boldsymbol{X}) = 0 \text{ or } L(\boldsymbol{\theta}^*, \boldsymbol{X}) = 0$$

  and

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle \neq \langle \boldsymbol{X}^*, \boldsymbol{X}^* \rangle \Rightarrow L(\boldsymbol{\theta}, \boldsymbol{X}) = 0 \text{ or } L(\boldsymbol{\theta}, \boldsymbol{X}^*) = 0$$

  where $\langle \cdot, \cdot \rangle$ denotes the quadratic variation
- Therefore unless $\boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta}) = \boldsymbol{\beta}(\boldsymbol{X}_t, \boldsymbol{\theta}^*)$, the acceptance probability contains the factor 0
- Hence the scheme becomes degenerate

Why the improvement?

- Under the innovation scheme as $m \to \infty$
- We have $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ which are consistent with $\boldsymbol{X}$ and $\boldsymbol{X}^*$ respectively
- Therefore in the acceptance probability the numerator and denominator differ in $\boldsymbol{\theta}$ and $\boldsymbol{X}$
- Thus a situation where the scheme becomes degenerate should not occur

Future work

- Extend these methods to data where we only have partial observations
- Examine the case where we have observations observed with error, typically Gaussian error
- Extend these methods to incorporate mixed effects

References

- 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. and Wilkinson, D. J. *Bayesian inference for nonlinear multivariate diffusion models observed with error.* Computational Statistics and Data Analysis, 52 (3) 1674-1693, 2008
- Wilkinson, D. J. and Golightly, A. *Markov chain Monte Carlo algorithms for SDE parameter estimation.* Learning and Inference in Computational Systems Biology, 2010
- Boys, R. J. Wilkinson, D. J. and Kirkwood, T. B. L. *Bayesian inference for a discretely observed stochastic kinetic model.* Statistics and Computing, 18 (2) 125-135, 2008
- Dargatz, C. *Bayesian Inference for Diffusion Processes with Applications in Life Sciences* PhD thesis, Ludwig–Maximilians–Universität München, 2010