Bayesian inference for discretely observed diffusion processes

Moritz Schauer
with Frank van der Meulen and Harry van Zanten

Delft University of Technology, University of Amsterdam

Van Dantzig Seminar

Estimating parameters of a discretely observed diffusion process

Diffusion process X

$$\mathrm{d}X_t = b_\theta(t, X_t)\,\mathrm{d}t + \sigma_\theta(t, X_t)\,\mathrm{d}W_t, \quad X_0 = u,$$

with transition densities $p(s, x; t, y)$
Discrete observations

$$X_{t_i} = x_i, \quad 0 = t_0 < t_1 < \cdots < t_n.$$

- Bayesian estimate for parameter θ with prior $\pi_0(\theta)$.
- Likelihood is intractable (product of transition densities)
- Continuous time likelihood known in closed form (Girsanov's theorem)

Computational approach

Data Augmentation (DA): Sample from the joint posterior of missing data and parameter.

1. Sample diffusion bridges conditional on $\{X_{t_i} = x_i\}$ and θ (this gives "complete", latent data);
2. Sample from θ conditional on the complete data.

Can use an accept/reject or Metropolis-Hastings step.

Rough outline:

- Simulation of diffusion bridges
- If unknown parameters are in the diffusion coefficient, DA does not work
- Example
- When and how to discretize

Computational approach

Data Augmentation (DA): Sample from the joint posterior of missing data and parameter.

1. Sample diffusion bridges conditional on $\{X_{t_i} = x_i\}$ and θ (this gives "complete", latent data);
2. Sample from θ conditional on the complete data.

Can use an accept/reject or Metropolis-Hastings step.

Rough outline:

- Simulation of diffusion bridges
- If unknown parameters are in the diffusion coefficient, DA does not work
- Example
- When and how to discretize

Computational approach

Data Augmentation (DA): Sample from the joint posterior of missing data and parameter.

1. Sample diffusion bridges conditional on $\{X_{t_i} = x_i\}$ and θ (this gives "complete", latent data);
2. Sample from θ conditional on the complete data.

Can use an accept/reject or Metropolis-Hastings step.

Rough outline:

- Simulation of diffusion bridges
- If unknown parameters are in the diffusion coefficient, DA does not work
- Example
- When and how to discretize

Computational approach

Data Augmentation (DA): Sample from the joint posterior of missing data and parameter.

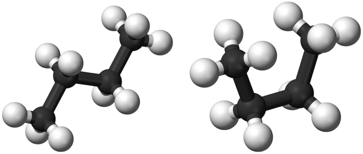
1. Sample diffusion bridges conditional on $\{X_{t_i} = x_i\}$ and θ (this gives "complete", latent data);
2. Sample from θ conditional on the complete data.

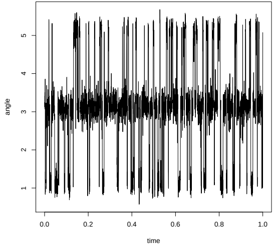
Can use an accept/reject or Metropolis-Hastings step.

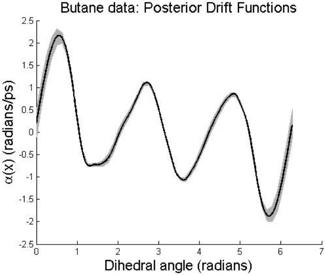
Rough outline:

- Simulation of diffusion bridges
- If unknown parameters are in the diffusion coefficient, DA does not work
- Example
- When and how to discretize

Examples: Butane dihedral angle, Pokern (2007)

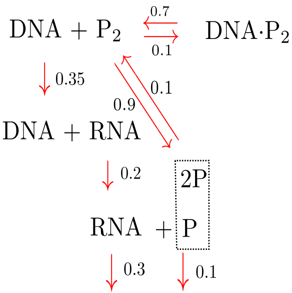


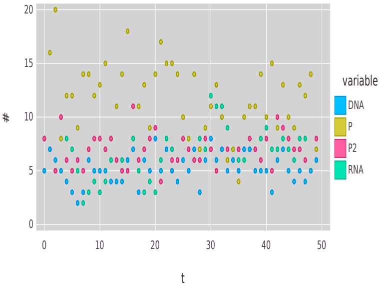




$$\mathrm{d}X_t = \sum_{i=1}^{J} \theta_i \psi_i(X_t)\,\mathrm{d}t + \mathrm{d}W_t$$

Chemical reaction network, Golightly and Wilkinson (2010)





$$dX_t = Sh_\theta(X_t)\,dt + S\,\mathrm{diag}(\sqrt{h_\theta(X_t)})\,dW_t$$

Intuition: Diffusion bridge

Two processes with equivalent distributions $\mathbb{P}$ and $\mathbb{W}$

- Diffusion process X with $\sigma \equiv 1$ starting in u
- Brownian motion W starting in u

Brownian motion W conditional on $W_T = v$: Brownian bridge.

The two conditional distributions $\mathbb{P}^\star$ and $\mathbb{W}^\star$ given $X_T = v$ resp. $W_T = v$ are equivalent

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}} = \frac{p(0,u;T,v)}{\phi(0,u;T,v)} \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{W}^\star}$$

with p and ϕ denoting the transition densities.

Works only if σ is constant. More general bridge proposals X° are needed,

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = C\Psi(X^\circ)$$

Intuition: Diffusion bridge

Two processes with equivalent distributions $\mathbb{P}$ and $\mathbb{W}$

- Diffusion process X with $\sigma \equiv 1$ starting in u
- Brownian motion W starting in u

Brownian motion W conditional on $W_T = v$: Brownian bridge.

The two conditional distributions $\mathbb{P}^\star$ and $\mathbb{W}^\star$ given $X_T = v$ resp. $W_T = v$ are equivalent

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}} = \frac{p(0, u; T, v)}{\phi(0, u; T, v)} \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{W}^\star}$$

with p and ϕ denoting the transition densities.

Works only if σ is constant. More general bridge proposals X° are needed,

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = C\Psi(X^\circ)$$

Intuition: Diffusion bridge

Two processes with equivalent distributions $\mathbb{P}$ and $\mathbb{W}$

- Diffusion process X with $\sigma \equiv 1$ starting in u
- Brownian motion W starting in u

Brownian motion W conditional on $W_T = v$: Brownian bridge.

The two conditional distributions $\mathbb{P}^\star$ and $\mathbb{W}^\star$ given $X_T = v$ resp. $W_T = v$ are equivalent

$$\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{W}} = \frac{p(0,u;T,v)}{\phi(0,u;T,v)} \frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{W}^\star}$$

with p and ϕ denoting the transition densities.

Works only if σ is constant. More general bridge proposals X° are needed,

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = C\Psi(X^\circ)$$

Diffusion bridges

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\star = b^\star(t, X_t^\star)\,\mathrm{d}t + \sigma(t, X_t^\star)\,\mathrm{d}W_t, \quad X_0^\star = u$$

with drift ($a = \sigma\sigma'$)

$$b^\star(t,x) = b(t,x) + a(t,x)\underbrace{\nabla_x \log p(t,x;T,v)}_{r(t,x;T,v)}.$$

- Delyon & Hu, Durham & Gallant: Proposals X° of the form

$$dX_t^\circ = \left(\lambda b(t, X_t^\circ) + \frac{v - X_t^\circ}{T - t}\right)\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u.$$

  $\lambda \in \{0, 1\}$.
- Beskos & Roberts: rejection sampling algorithm for obtaining bridges without discretisation error.

Diffusion bridges

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\star = b^\star(t, X_t^\star)\,\mathrm{d}t + \sigma(t, X_t^\star)\,\mathrm{d}W_t, \quad X_0^\star = u$$

with drift ($a = \sigma\sigma'$)

$$b^\star(t, x) = b(t, x) + a(t, x)\underbrace{\nabla_x \log p(t, x; T, v)}_{r(t, x; T, v)}.$$

- Delyon & Hu, Durham & Gallant: Proposals X° of the form

$$dX_t^\circ = \left(\lambda b(t, X_t^\circ) + \frac{v - X_t^\circ}{T - t}\right)\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u.$$

  $\lambda \in \{0, 1\}$.
- Beskos & Roberts: rejection sampling algorithm for obtaining bridges without discretisation error.

Diffusion bridges

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\star = b^\star(t, X_t^\star)\,\mathrm{d}t + \sigma(t, X_t^\star)\,\mathrm{d}W_t, \quad X_0^\star = u$$

with drift ($a = \sigma\sigma'$)

$$b^\star(t, x) = b(t, x) + a(t, x)\underbrace{\nabla_x \log p(t, x; T, v)}_{r(t,x;T,v)}.$$

- Delyon & Hu, Durham & Gallant: Proposals X° of the form

$$dX_t^\circ = \left(\lambda b(t, X_t^\circ) + \frac{v - X_t^\circ}{T - t}\right)\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u.$$

  $\lambda \in \{0, 1\}$.
- Beskos & Roberts: rejection sampling algorithm for obtaining bridges without discretisation error.

Diffusion bridge proposals

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\star = b^\star(t, X_t^\star)\,\mathrm{d}t + \sigma(t, X_t^\star)\,\mathrm{d}W_t, \quad X_0^\star = u$$

with drift

$$b^\star(t, x) = b(t, x) + a(t, x)\underbrace{\nabla_x \log p(t, x; T, v)}_{r(t, x; T, v)}.$$

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\circ = b^\circ(t, X_t^\circ)\,\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u$$

with drift

$$b^\circ(t, x) = b(t, x) + a(t, x)\underbrace{\nabla_x \log \tilde{p}(t, x; T, v)}_{\tilde{r}(t, x; T, v)}.$$

Take $\tilde{p}$ the transition density of

$$\mathrm{d}\tilde{X}_t = \left(\tilde{\beta}(t) + \tilde{B}(t)\tilde{X}_t\right)\,\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}W_t.$$

If $\tilde{a}(T) = a(T, v)$ (and a few more conditions), then

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{p}(0, u; T, v)}{p(0, u; T, v)}\Psi(X^\circ)$$

Diffusion bridge proposals

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\circ = b^\circ(t, X_t^\circ)\,\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u$$

with drift

$$b^\circ(t, x) = b(t, x) + a(t, x)\underbrace{\nabla_x \log \tilde{p}(t, x; T, v)}_{\tilde{r}(t, x; T, v)}.$$

Take $\tilde{p}$ the transition density of

$$\mathrm{d}\tilde{X}_t = \left(\tilde{\beta}(t) + \tilde{B}(t)\tilde{X}_t\right)\,\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}W_t.$$

If $\tilde{a}(T) = a(T, v)$ (and a few more conditions), then

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{p}(0, u; T, v)}{p(0, u; T, v)}\Psi(X^\circ)$$

where Ψ is tractable.

Diffusion bridge proposals

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\circ = b^\circ(t, X_t^\circ)\,\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u$$

with drift

$$b^\circ(t,x) = b(t,x) + a(t,x)\underbrace{\nabla_x \log \tilde{p}(t,x;T,v)}_{\tilde{r}(t,x;T,v)}.$$

Take $\tilde{p}$ the transition density of

$$\mathrm{d}\tilde{X}_t = \left(\tilde{\beta}(t) + \tilde{B}(t)\tilde{X}_t\right)\,\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}W_t.$$

If $\tilde{a}(T) = a(T, v)$ (and a few more conditions), then

$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{p}(0,u;T,v)}{p(0,u;T,v)}\Psi(X^\circ)$$

where Ψ is tractable.

Diffusion bridge proposals

Bridge from $(0, u)$ to (T, v)

$$\mathrm{d}X_t^\circ = b^\circ(t, X_t^\circ)\,\mathrm{d}t + \sigma(t, X_t^\circ)\,\mathrm{d}W_t, \quad X_0^\circ = u$$

with drift

$$b^\circ(t,x) = b(t,x) + a(t,x)\underbrace{\nabla_x \log \tilde{p}(t,x;T,v)}_{\tilde{r}(t,x;T,v)}.$$

Take $\tilde{p}$ the transition density of

$$\mathrm{d}\tilde{X}_t = \left(\tilde{\beta}(t) + \tilde{B}(t)\tilde{X}_t\right)\,\mathrm{d}t + \tilde{\sigma}(t)\,\mathrm{d}W_t.$$

If $\tilde{a}(T) = a(T, v)$ (and a few more conditions), then

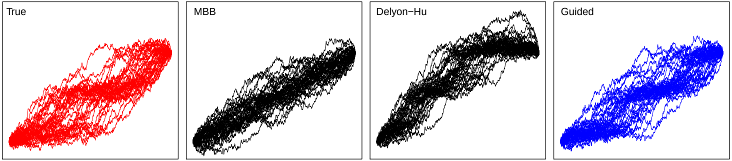
$$\frac{\mathrm{d}\mathbb{P}^\star}{\mathrm{d}\mathbb{P}^\circ}(X^\circ) = \frac{\tilde{p}(0,u;T,v)}{p(0,u;T,v)}\Psi(X^\circ)$$

where Ψ is tractable.

An example

Example: Simulate X given that $X_0 = 0$ and $X_1 = \pi/2$.

$$\mathrm{d}X_t = (2 - 2\sin(8X_t))\,\mathrm{d}t + \tfrac{1}{2}\,\mathrm{d}W_t$$



Guided proposal from

$$\mathrm{d}\tilde{X}_t = 1.34\,\mathrm{d}t + \frac{1}{2}\,\mathrm{d}W_t.$$

yielding

$$\mathrm{d}X_t^\circ = \left(2 - 2\sin(8X_t^\circ) + \frac{\pi/2 - X_t^\circ}{1-t} - 1.34\right)\mathrm{d}t + \frac{1}{2}\,\mathrm{d}W_t, \quad X_0^\circ = 0.$$

Finding good proposals

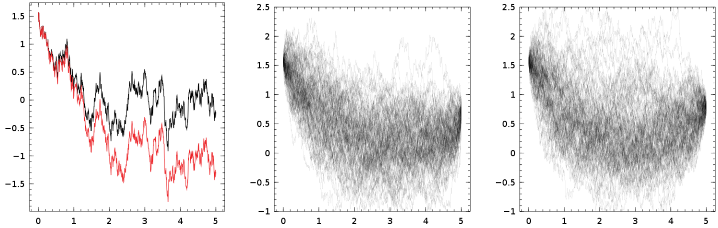
- Cross entropy method (previous example)
- Local linearizations (chemical reaction network example)
- Substituting space dependence for time dependence (next slide)

Substituting space dependence for time dependence

$$\mathrm{d}X_t = -\sin(X_t)\,\mathrm{d}t, \quad X_0 = \pi/2$$
$$\mathrm{d}\tilde{X}_t = -\operatorname{sech}(t)\,\mathrm{d}t, \quad \tilde{X}_0 = \pi/2$$

Substituting space dependence for time dependence

$$\mathrm{d}X_t = -\sin(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0 = \pi/2$$
$$\mathrm{d}\tilde{X}_t = -\operatorname{sech}(t)\,\mathrm{d}t + \mathrm{d}W_t, \quad \tilde{X}_0 = \pi/2$$



Sample of X and $\tilde{X}$ (black, red) X° $X^\star$.

Updating bridges

- Use a Metropolis-Hastings step with independent proposals from X°.
- Assume one bridge with $X^\circ_0 = u$, $X^\circ_T = v$.
- Sample proposal process X°. Accept proposal X° with probability $\min(1, A)$

$$A = \frac{\Psi(X^\circ)}{\Psi(X^\star)}$$

else retain current process $X^\star$.

Updating bridges

- Use a Metropolis-Hastings step with independent proposals from X°.
- Assume one bridge with $X^\circ_0 = u$, $X^\circ_T = v$.
- Sample proposal process X°. Accept proposal X° with probability $\min(1, A)$
$$A = \frac{\Psi(X^\circ)}{\Psi(X^\star)}$$
else retain current process $X^\star$.

Parameters in the diffusion coefficient: DA does not work

Toy example from Roberts and Stramer (2001).

Consider the diffusion generated by the SDE

$$\mathrm{d}X_t = \theta\,\mathrm{d}W_t, \qquad X_0 = 0$$

X_1 is observed. θ unknown.

Intuitive argument why DA fails:

- Initialize θ by θ_0 and simulate a bridge $X^\star$ in continuous time.
- $X^\star$ has quadratic variation θ_0, so any new iterate for θ must be θ_0.

Parameters in the diffusion coefficient: DA does not work

Toy example from ROBERTS AND STRAMER (2001).

Consider the diffusion generated by the SDE

$$\mathrm{d}X_t = \theta \, \mathrm{d}W_t, \qquad X_0 = 0$$

X_1 is observed. θ unknown.

Intuitive argument why DA fails:

- Initialize θ by θ_0 and simulate a bridge $X^\star$ in continuous time.
- $X^\star$ has quadratic variation θ_0, so any new iterate for θ must be θ_0.

Parameters in the diffusion coefficient: DA does not work

Toy example from ROBERTS AND STRAMER (2001).

Consider the diffusion generated by the SDE

$$\mathrm{d}X_t = \theta \, \mathrm{d}W_t, \qquad X_0 = 0$$

X_1 is observed. θ unknown.

Intuitive argument why DA fails:

- Initialize θ by θ_0 and simulate a bridge $X^\star$ in continuous time.
- $X^\star$ has quadratic variation θ_0, so any new iterate for θ must be θ_0.

Reparametrisation

- Let Z° be a Wiener process. We have

$$\mathrm{d}X_t^\circ = (b_\theta + a_\theta \tilde{r}_\theta)(t, X_t^\circ)\,\mathrm{d}t + \sigma_\theta(t, X_t^\circ)\,\mathrm{d}Z_t^\circ.$$

  We write

$$X^\circ = g(\theta, Z^\circ).$$

- Similarly

$$X^\star = g(\theta, Z^\star)$$

  by taking

$$Z_t^\star = W_t + \int_0^t \sigma_\theta'(s, X_s^\star)\,(r_\theta(t, X_t^\star) - \tilde{r}_\theta(s, X_s^\star))\,\mathrm{d}s$$

Reparametrisation

- Let Z° be a Wiener process. We have

$$\mathrm{d}X^\circ_t = (b_\theta + a_\theta \tilde{r}_\theta)(t, X^\circ_t)\,\mathrm{d}t + \sigma_\theta(t, X^\circ_t)\,\mathrm{d}Z^\circ_t.$$

  We write

$$X^\circ = g(\theta, Z^\circ).$$

- Similarly

$$X^\star = g(\theta, Z^\star)$$

  by taking

$$Z^\star_t = W_t + \int_0^t \sigma'_\theta(s, X^\star_s)\left(r_\theta(t, X^\star_t) - \tilde{r}_\theta(s, X^\star_s)\right)\,\mathrm{d}s$$

Reparametrisation

Key idea: Sample from θ conditional on $(X_0 = u, X_T = v, Z^\star)$ instead of θ conditional on $(X_0 = u, X_T = v, X^\star)$.

Notation:

Process	$Z^\star \xrightarrow{g(\theta,\cdot)} X^\star$		$Z^\circ \xrightarrow{g(\theta,\cdot)} X^\circ$	
Law	$\mathbb{Q}^\star_\theta$	$\mathbb{P}^\star_\theta$	$\mathbb{Q}^\circ$	$\mathbb{P}^\circ_\theta$

Metropolis-Hastings step: Propose a value θ° from some proposal distribution $q(\cdot \mid \theta)$ and accept the proposal with probability $\min(1, A)$, where

$$A = \underbrace{\frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}}_{\text{prior ratio}} \underbrace{\frac{p_{\theta^\circ}(0,u;T,v)}{p_\theta(0,u;T,v)} \frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star)}_{\text{likelihood ratio}} \underbrace{\frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)}}_{\text{proposal ratio}}.$$

Reparametrisation

Key idea: Sample from θ conditional on $(X_0 = u, X_T = v, Z^\star)$ instead of θ conditional on $(X_0 = u, X_T = v, X^\star)$.

Notation:

Process	$Z^\star \xrightarrow{g(\theta,\cdot)}$	$X^\star$	$Z^\circ \xrightarrow{g(\theta,\cdot)}$	X°
Law	$\mathbb{Q}^\star_\theta$	$\mathbb{P}^\star_\theta$	$\mathbb{Q}^\circ$	$\mathbb{P}^\circ_\theta$

Metropolis-Hastings step: Propose a value θ° from some proposal distribution $q(\cdot \mid \theta)$ and accept the proposal with probability $\min(1, A)$, where

$$A = \underbrace{\frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}}_{\text{prior ratio}} \underbrace{\frac{p_{\theta^\circ}(0,u;T,v)}{p_\theta(0,u;T,v)} \frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star)}_{\text{likelihood ratio}} \underbrace{\frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)}}_{\text{proposal ratio}} .$$

Reparametrisation

Key idea: Sample from θ conditional on $(X_0 = u, X_T = v, Z^\star)$ instead of θ conditional on $(X_0 = u, X_T = v, X^\star)$.

Notation:

Process	$Z^\star \xrightarrow{g(\theta,\cdot)} X^\star$		$Z^\circ \xrightarrow{g(\theta,\cdot)} X^\circ$	
Law	$\mathbb{Q}^\star_\theta$	$\mathbb{P}^\star_\theta$	$\mathbb{Q}^\circ$	$\mathbb{P}^\circ_\theta$

Metropolis-Hastings step: Propose a value θ° from some proposal distribution $q(\cdot \mid \theta)$ and accept the proposal with probability $\min(1, A)$, where

$$A = \underbrace{\frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}}_{\text{prior ratio}} \underbrace{\frac{p_{\theta^\circ}(0,u;T,v)}{p_\theta(0,u;T,v)} \frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star)}_{\text{likelihood ratio}} \underbrace{\frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)}}_{\text{proposal ratio}} .$$

Reparametrisation

Metropolis-Hastings step: Propose a value θ° from some proposal distribution $q(\cdot \mid \theta)$ and accept the proposal with probability $\min(1, A)$, where

$$A = \underbrace{\frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}}_{\text{prior ratio}} \underbrace{\frac{p_{\theta^\circ}(0,u;T,v)}{p_\theta(0,u;T,v)} \frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star)}_{\text{likelihood ratio}} \underbrace{\frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)}}_{\text{proposal ratio}} .$$

Using absolute continuity results of $\mathbb{P}^\star$ wrt $\mathbb{P}^\circ$, we get

$$\frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star) = \frac{p_\theta(0,u;T,v)}{p_{\theta^\circ}(0,u;T,v)} \frac{\tilde{p}_{\theta^\circ}(0,u;T,v)}{\tilde{p}_\theta(0,u;T,v)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z^\star))}{\Psi_\theta(g(\theta, Z^\star))}.$$

Reparametrisation

Metropolis-Hastings step: Propose a value θ° from some proposal distribution $q(\cdot \mid \theta)$ and accept the proposal with probability $\min(1, A)$, where

$$A = \underbrace{\frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}}_{\text{prior ratio}} \underbrace{\frac{p_{\theta^\circ}(0,u;T,v)}{p_\theta(0,u;T,v)} \frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star)}_{\text{likelihood ratio}} \underbrace{\frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)}}_{\text{proposal ratio}} .$$

Using absolute continuity results of $\mathbb{P}^\star$ wrt $\mathbb{P}^\circ$, we get

$$\frac{\mathrm{d}\mathbb{Q}^\star_{\theta^\circ}}{\mathrm{d}\mathbb{Q}^\star_\theta}(Z^\star) = \frac{p_\theta(0,u;T,v)}{p_{\theta^\circ}(0,u;T,v)} \frac{\tilde{p}_{\theta^\circ}(0,u;T,v)}{\tilde{p}_\theta(0,u;T,v)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z^\star))}{\Psi_\theta(g(\theta, Z^\star))}.$$

Algorithm

1. **Update** $Z^\star \mid (\theta, X_{t_i} = x_{i,1\le i\le n})$. Independently, for $1 \le i \le n$ do
   1.1 Sample a Wiener process Z_i°.
   1.2 Sample $U \sim \mathcal{U}(0,1)$. Compute

$$A_1 = \frac{\Psi_\theta(g(\theta, Z_i^\circ))}{\Psi_\theta(g(\theta, Z_i^\star))}.$$

Set

$$Z_i^\star := \begin{cases} Z_i^\circ & \text{if} \quad U \le A_1 \\ Z_i^\star & \text{if} \quad U > A_1 \end{cases}.$$

2. **Update** $\theta \mid (Z^\star, X_{t_i} = x_{i,1\le i\le n})$.
   2.1 Sample $\theta^\circ \sim q(\cdot \mid \theta)$.
   2.2 Sample $U \sim \mathcal{U}(0,1)$. Compute

$$A_2 = \prod_{i=1}^{n} \frac{\tilde{p}_{\theta^\circ}(t_{i-1}, x_{i-1}; t_i, x_i)}{\tilde{p}_\theta(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z_i^\star))}{\Psi_\theta(g(\theta, Z_i^\star))} \frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)} \frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}$$

Set

$$\theta := \begin{cases} \theta^\circ & \text{if} \quad U \le A_2 \\ \theta & \text{if} \quad U > A_2 \end{cases}.$$

Repeat steps (1) and (2).

Algorithm

1. **Update** $Z^\star \mid (\theta, X_{t_i} = x_{i,1 \le i \le n})$**.** Independently, for $1 \le i \le n$ do
   1.1 Sample a Wiener process Z_i°.
   1.2 Sample $U \sim \mathcal{U}(0,1)$. Compute

$$A_1 = \frac{\Psi_\theta(g(\theta, Z_i^\circ))}{\Psi_\theta(g(\theta, Z_i^\star))}.$$

   Set

$$Z_i^\star := \begin{cases} Z_i^\circ & \text{if} \quad U \le A_1 \\ Z_i^\star & \text{if} \quad U > A_1 \end{cases}.$$

2. **Update** $\theta \mid (Z^\star, X_{t_i} = x_{i,1 \le i \le n})$**.**
   2.1 Sample $\theta^\circ \sim q(\cdot \mid \theta)$.
   2.2 Sample $U \sim \mathcal{U}(0,1)$. Compute

$$A_2 = \prod_{i=1}^{n} \frac{\tilde{p}_{\theta^\circ}(t_{i-1}, x_{i-1}; t_i, x_i)}{\tilde{p}_\theta(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z_i^\star))}{\Psi_\theta(g(\theta, Z_i^\star))} \frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)} \frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}$$

   Set

$$\theta := \begin{cases} \theta^\circ & \text{if} \quad U \le A_2 \\ \theta & \text{if} \quad U > A_2 \end{cases}.$$

Repeat steps (1) and (2).

Algorithm

1. **Update** $Z^\star \mid (\theta, X_{t_i} = x_{i,1\le i\le n})$**.** Independently, for $1 \le i \le n$ do
   1.1 Sample a Wiener process Z_i°.
   1.2 Sample $U \sim \mathcal{U}(0,1)$. Compute
   $$A_1 = \frac{\Psi_\theta(g(\theta, Z_i^\circ))}{\Psi_\theta(g(\theta, Z_i^\star))}.$$
   Set
   $$Z_i^\star := \begin{cases} Z_i^\circ & \text{if} \quad U \le A_1 \\ Z_i^\star & \text{if} \quad U > A_1 \end{cases}.$$
2. **Update** $\theta \mid (Z^\star, X_{t_i} = x_{i,1\le i\le n})$**.**
   2.1 Sample $\theta^\circ \sim q(\cdot \mid \theta)$.
   2.2 Sample $U \sim \mathcal{U}(0,1)$. Compute
   $$A_2 = \prod_{i=1}^{n} \frac{\tilde{p}_{\theta^\circ}(t_{i-1}, x_{i-1}; t_i, x_i)}{\tilde{p}_\theta(t_{i-1}, x_{i-1}; t_i, x_i)} \frac{\Psi_{\theta^\circ}(g(\theta^\circ, Z_i^\star))}{\Psi_\theta(g(\theta, Z_i^\star))} \frac{q(\theta \mid \theta^\circ)}{q(\theta^\circ \mid \theta)} \frac{\pi_0(\theta^\circ)}{\pi_0(\theta)}$$
   Set
   $$\theta := \begin{cases} \theta^\circ & \text{if} \quad U \le A_2 \\ \theta & \text{if} \quad U > A_2 \end{cases}.$$

Repeat steps (1) and (2).

Numerical example: Data

Prokaryotic auto-regulation example, Golightly and Wilkinson (2010)

Markov chain modelling quantities of $(\mathrm{RNA}, \mathrm{P}, \mathrm{P}_2, \mathrm{DNA})$ at integer times

Numerical example: SDE

Chemical Langevin equation. Diffusion approximation of the Markov chain

$$\mathrm{d}X_t = Sh_\theta(X_t)\,\mathrm{d}t + S\,\mathrm{diag}(\sqrt{h_\theta(X_t)})\,\mathrm{d}W_t$$

driven by a $\mathbb{R}^8$-valued Brownian motion, where

$$S = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 2 & 0 & -1 \\ -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$h_\theta(x) = \mathrm{diag}(\theta) \cdot [x_3 x_4, K - x_4, x_4, x_1, \tfrac{1}{2} x_2(x_2 - 1), x_3, x_1, x_2]'$$

Numerical example: Proposals

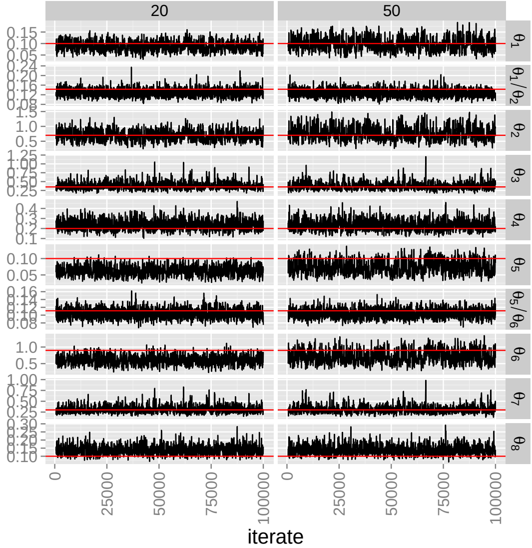
Linearization of drift

$\bar{h}_\theta(x) = \text{diag}(\theta) \cdot [c_1 + \lambda_1 x_3 + \gamma_1 x_4, K - x_4, x_4, x_1, c_2 + \lambda_2 x_2, x_3, x_1, x_2]'$

Choose $\tilde{B}_\theta$ and $\tilde{\beta}_\theta$ such that

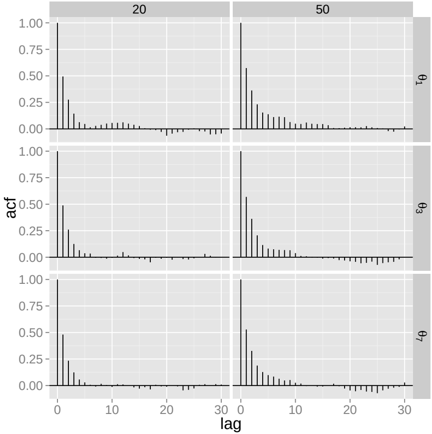
$$\tilde{B}_\theta x + \tilde{\beta}_\theta = S\bar{h}_\theta(x)$$

Numerical example: Results



Iterates of the MCMC chain for $m = 20, 50$ interpolated points

Numerical example: Results



ACF plots of the thinned samples of θ_1, θ_3, θ_7 (taking every 50th iterate after burn in).

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Summary

Data augmentation for discretely observed diffusion processes

Non-constant and unknown σ: two challenges

- Finding good proposals for the conditional process
- Overcoming dependence between missing data and parameter

Both can be addressed using guided proposals

- Proposal bridges which take drift into account
- Guided proposals provide a natural reparametrization to decouple parameter and latent path

Remarks

Overcoming singularities in the drift

Scaling and time change

$$U_s = \mathrm{e}^{s/2}(v - X^\circ_{T(1-\mathrm{e}^{-s})})$$

If the target is a Brownian motion, the proposal process U discretized on an equidistant grid and simulated using vanilla Euler scheme coincides with the scaled and time changed Brownian bridge (up to a small error)

Remarks

Determining $\tilde{B}$ and $\tilde{\beta}$

- Sometimes there a natural linearizations of the drift
- Adaptive proposals minimizing cross-entropy.