Particle Markov Chain Monte Carlo

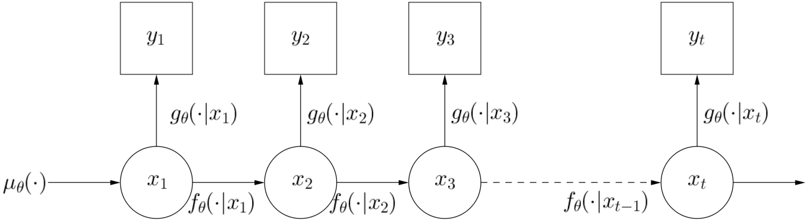
Ella Kaye   Nathan Cunningham   Kaspar Märtens

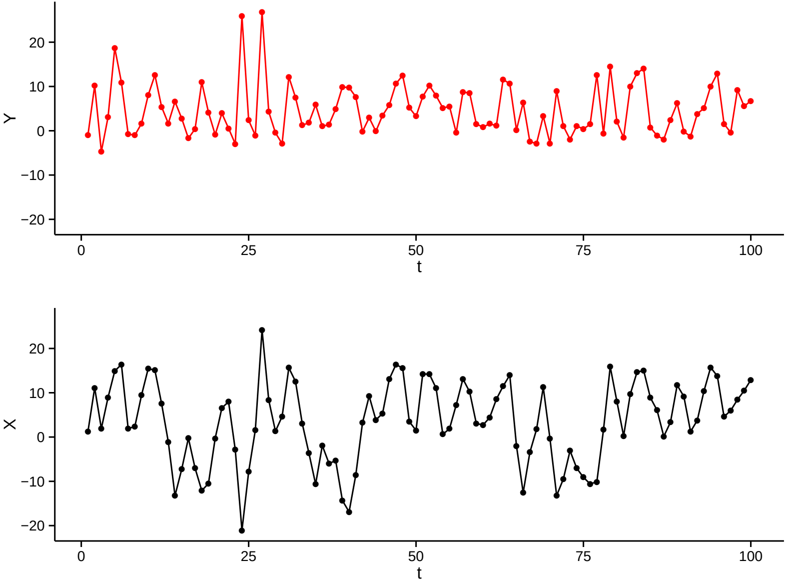
OxWaSP

20th November 2015

Particle Markov Chain Monte Carlo

- Combination of SMC steps within MCMC updates
- Takes advantage of strengths of both
- Particularly useful in state-space models


y_1
y_2
y_3
y_t
$g_\theta(\cdot|x_1)$
$g_\theta(\cdot|x_2)$
$g_\theta(\cdot|x_3)$
$g_\theta(\cdot|x_t)$
$\mu_\theta(\cdot)$
x_1
x_2
x_3
x_t
$f_\theta(\cdot|x_1)$
$f_\theta(\cdot|x_2)$
$f_\theta(\cdot|x_{t-1})$


20
10
0
−10
−20
Y
0
25
50
75
100
t
20
10
0
−10
−20
X
0
25
50
75
100
t

Sequential Monte Carlo (SMC)

Aim: approximate distribution $p(x_{1:t}|y_{1:t})$

- Standard MCMC methods are inefficient due to high dimensionality and dependence structure
- Idea of SMC: approximate $p(x_{1:t}|y_{1:t})$ sequentially for $t \geq 1$
  - First approximate $p(x_1|y_1)$ and $p(y_1)$
  - Then $p(x_{1:2}|y_{1:2})$ and $p(y_{1:2})$ and so on.

SMC algorithm

At time $t = 1$:

(a) Sample $X_1^k \sim q_\theta(\cdot|y_1)$ for all $k = 1, \dots, N$

(b) Compute and normalise the weights

$$w_1(X_1^k) := \frac{p_\theta(X_1^k, y_1)}{q_\theta(X_1^k|y_1)} = \frac{\mu_\theta(X_1^k) g_\theta(y_1|X_1^k)}{q_\theta(X_1^k|y_1)}, \tag{1}$$

$$W_1^k := \frac{w_1(X_1^k)}{\sum_{m=1}^N w_1(X_1^m)}.$$

SMC algorithm

At time $t = 2, \ldots, T$:

(a) Obtain $\bar{X}_{t-1} = (\bar{X}_{t-1}^1, \ldots, \bar{X}_{t-1}^N)$ by resampling $X_{t-1} = (X_{t-1}^1, \ldots, X_{t-1}^N)$ given W_{t-1}.

(b) Sample $X_t^k \sim q_\theta(\cdot | y_t, \bar{X}_{t-1})$ and set $X_{1:t}^k := (\bar{X}_{1:t-1}, X_t^k)$.

(c) Compute and normalise the weights

$$w_t(X_{1:t}^k) = \frac{f_\theta(X_t^k) g_\theta(y_t | \bar{X}_{t-1})}{q_\theta(X_t^k | y_t, \bar{X}_{t-1})}, \tag{2}$$

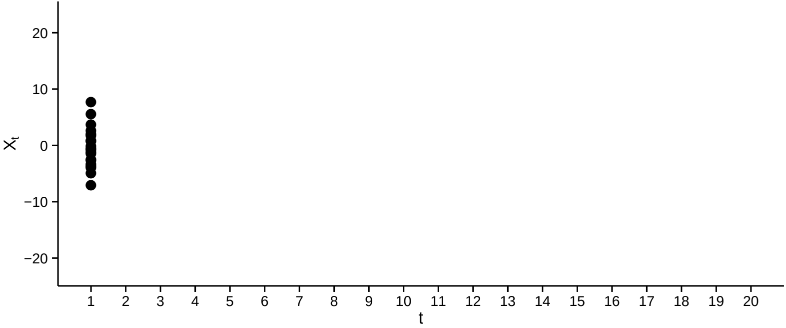
$$W_t^k := \frac{w_t(X_t^k)}{\sum_{m=1}^N w_t(X_t^m)}.$$

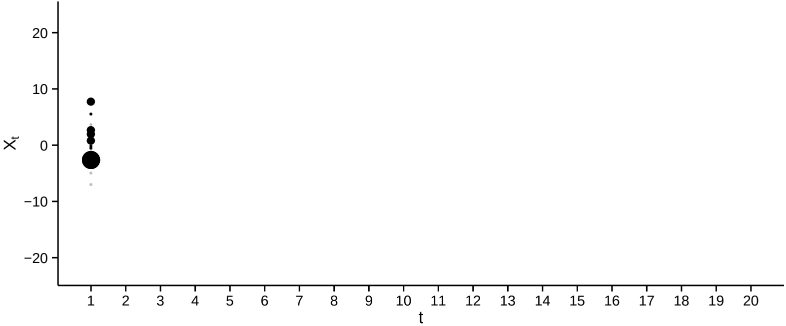
The SMC algorithm provides

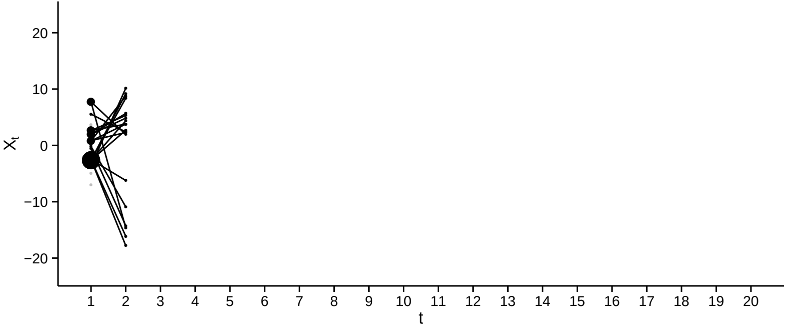
- an approximation of the joint posterior density at time T, $p_\theta(x_{1:t}|y_{1:t})$

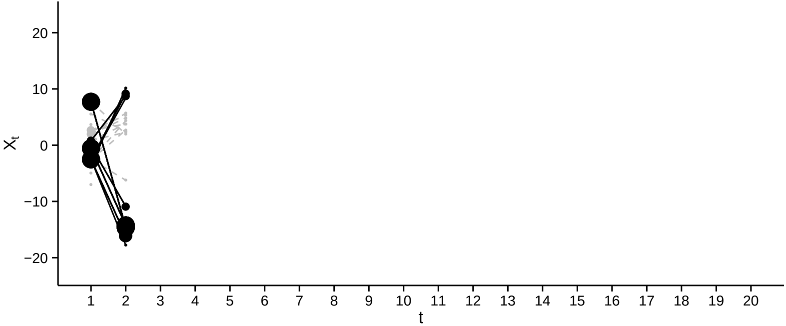
$$\hat{p}_\theta(\mathrm{d}x_{1:t}|y_{1:t}) := \sum_{k=1}^{N} W_T^k \delta_{X_{1:T}^k}(\mathrm{d}x_{1:T}), \tag{3}$$

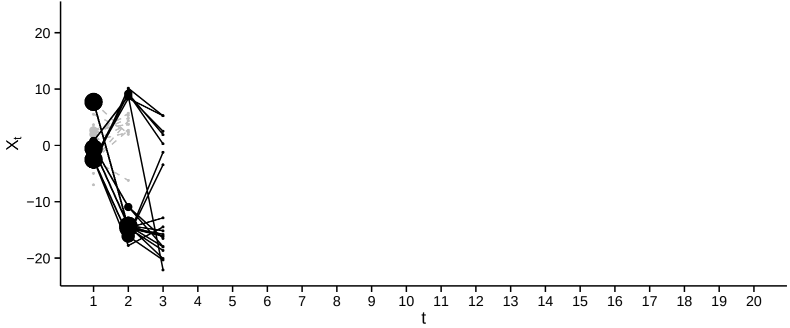
- an estimate of the marginal likelihood $p_\theta(y_{1:T})$

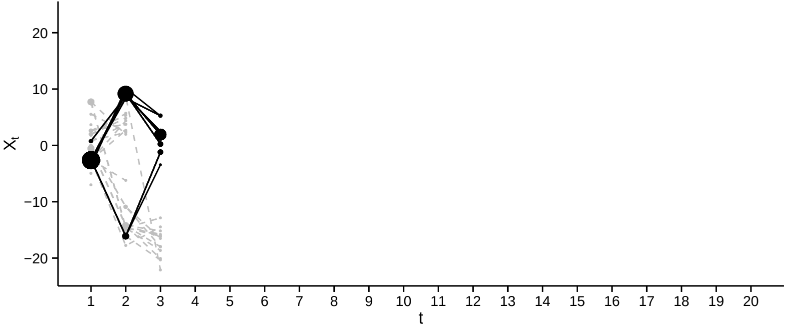

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

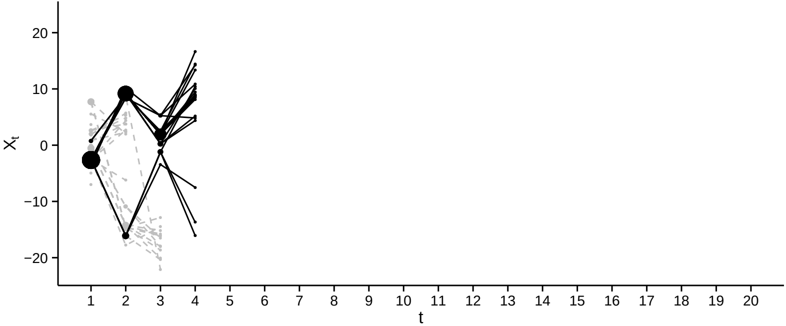

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

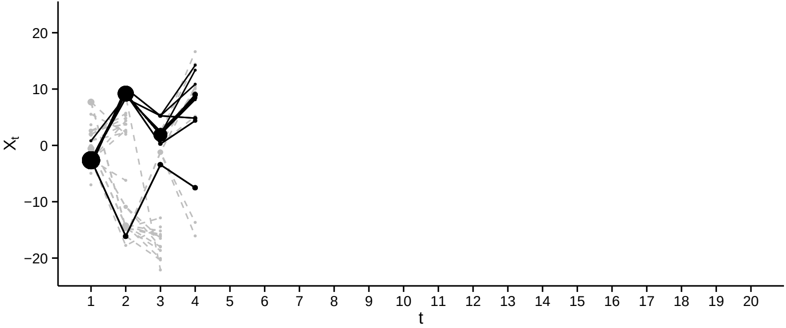

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

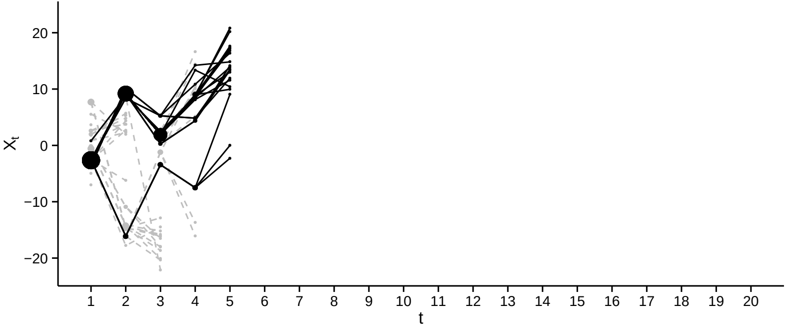

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

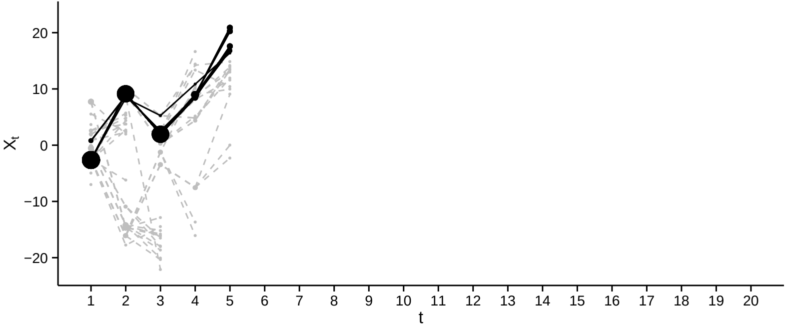

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

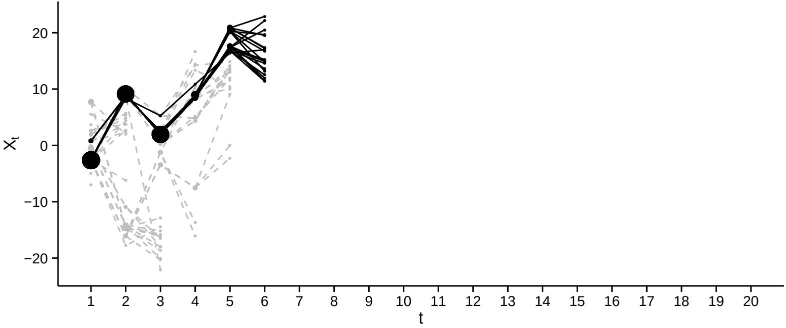

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

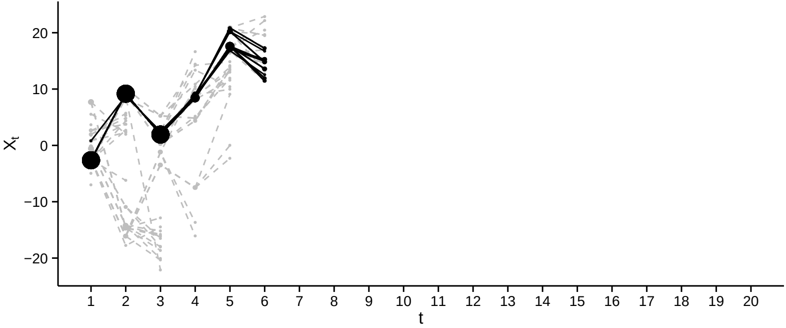

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

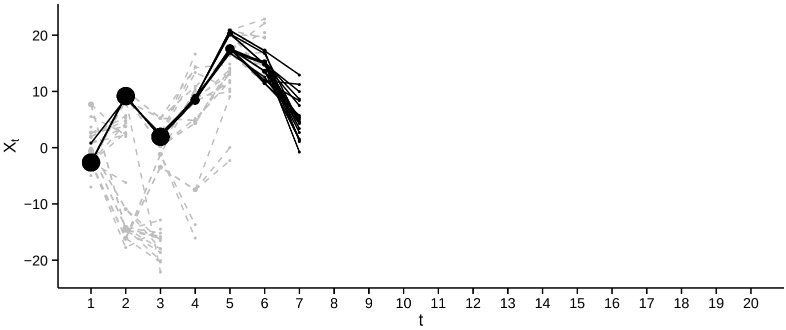

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

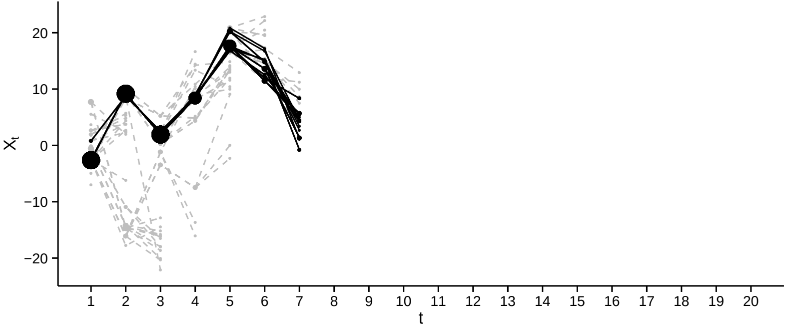

X_t
20
10
0
−10
−20
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
t

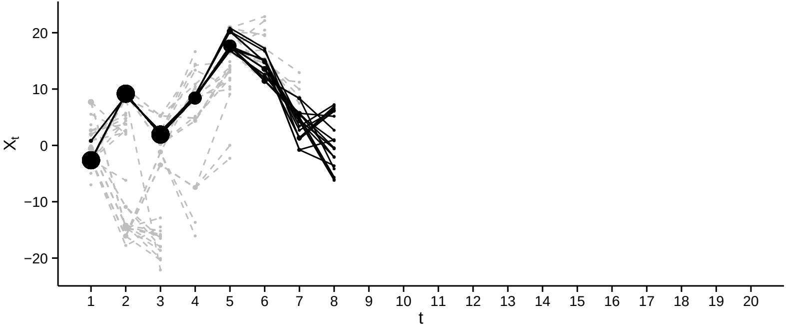

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

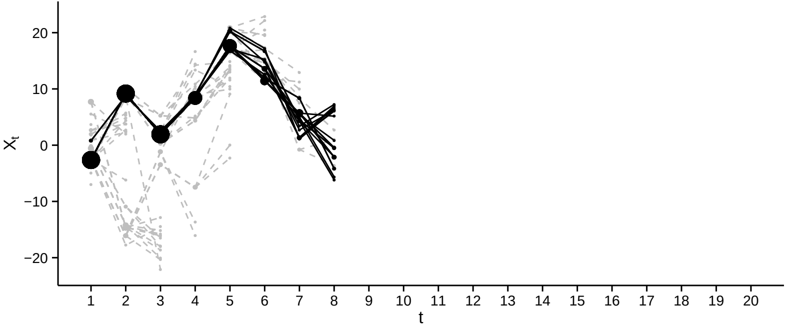

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

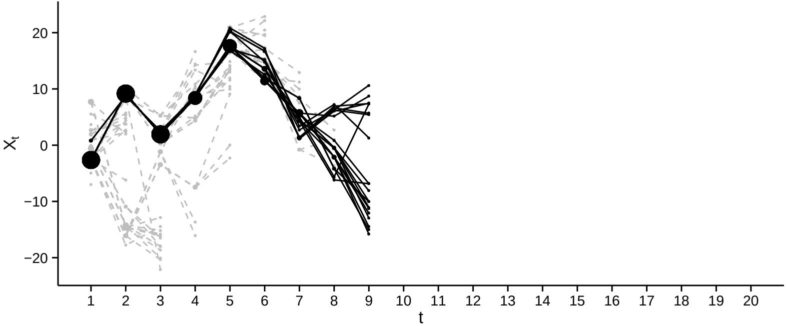

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

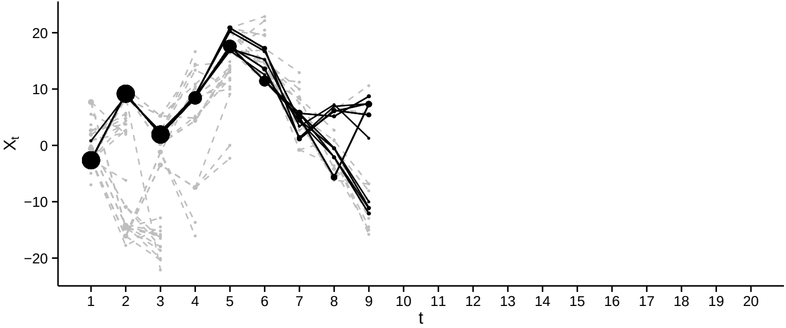

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

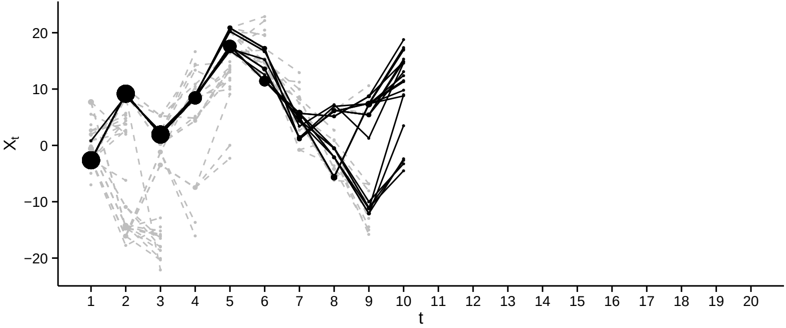

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

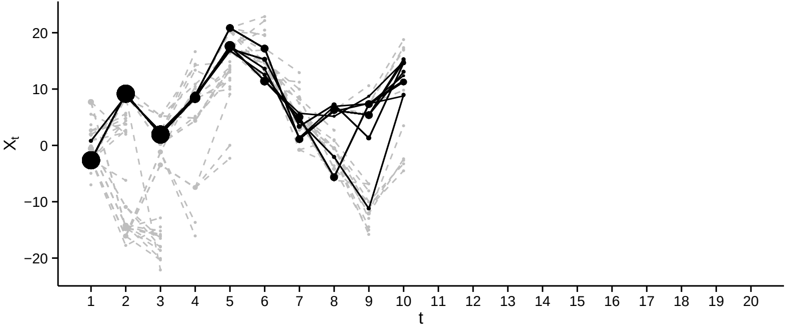

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

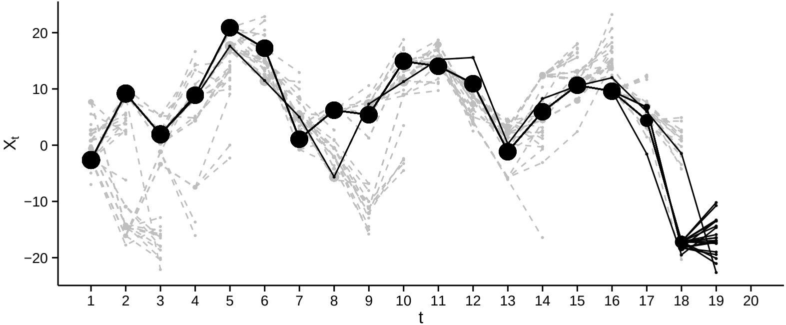

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

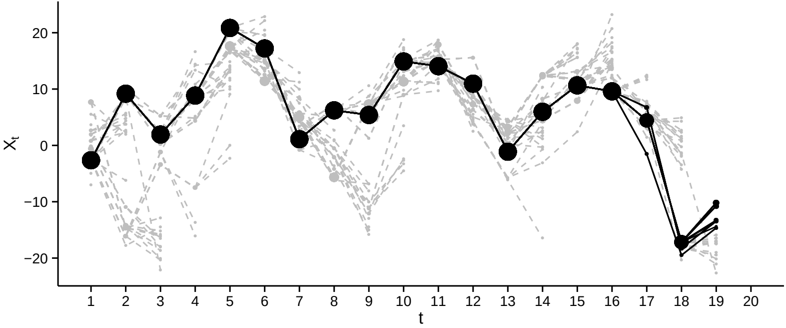

20
10
0
−10
−20
X_t
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
t

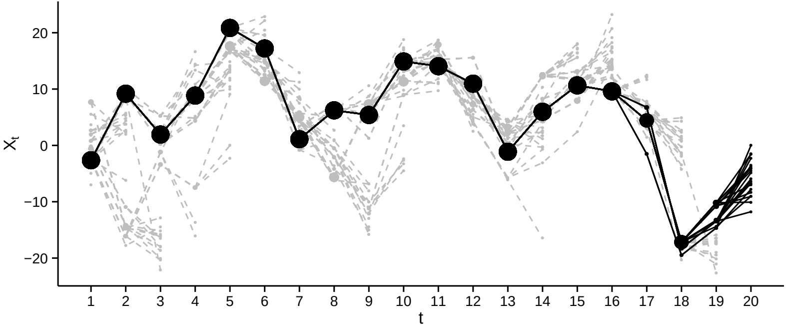

20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t


20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t


20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t


20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t


20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t


20
10
0
−10
−20
X_t
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
t

Particle MCMC

- Combination of standard MCMC and SMC
- New samples proposed using SMC
- Accept proposal according to particle Metropolis-Hastings probability

Particle MCMC

- Combination of standard MCMC and SMC
- New samples proposed using SMC
- Accept proposal according to particle Metropolis-Hastings probability

We consider two approaches:

- Particle Independent Metropolis-Hastings
- Particle Marginal Metropolis-Hastings

Particle Independent Metropolis-Hastings

Step 1: initialisation, $i = 0$:

(a) Run an SMC algorithm targeting $p_\theta(x_{1:T}|y_{1:T})$

(b) Sample $X_{1:T}(0) \sim \hat{p}_\theta(\cdot|y_{1:T})$ and let $\hat{p}_\theta(y_{1:T})(0)$ denote the corresponding marginal likelihood estimate.

Particle Independent Metropolis-Hastings

Step 1: initialisation, $i = 0$:

(a) Run an SMC algorithm targeting $p_\theta(x_{1:T}|y_{1:T})$

(b) Sample $X_{1:T}(0) \sim \hat{p}_\theta(\cdot|y_{1:T})$ and let $\hat{p}_\theta(y_{1:T})(0)$ denote the corresponding marginal likelihood estimate.

Step 2: for iteration $i \geq 1$:

(a) Run an SMC algorithm targeting $p_\theta(x_{1:T}|y_{1:T})$

(b) Sample $X^*_{1:T} \sim \hat{p}_\theta(\cdot|y_{1:T})$ and let $\hat{p}_\theta(y_{1:T})^*$ denote the corresponding marginal likelihood estimate.

(c) Accept update with probability

$$1 \wedge \frac{\hat{p}_\theta(y_{1:T})^*}{\hat{p}_\theta(y_{1:T})(i-1)} \tag{4}$$

Particle Marginal Metropolis-Hastings

Step 1: initialisation, $i = 0$, arbitrarily set $\theta(0)$

(a) Run an SMC algorithm targeting $p_{\theta(0)}(x_{1:T}|y_{1:T})$

(b) Sample $X_{1:T}(0) \sim \hat{p}_{\theta(0)}(\cdot|y_{1:T})$ and let $\hat{p}_{\theta(0)}(y_{1:T})(0)$ denote the corresponding marginal likelihood estimate.

Step 2: for iteration $i \geq 1$:

(a) Sample $\theta^* \sim q \cdot |\theta(i-1)$

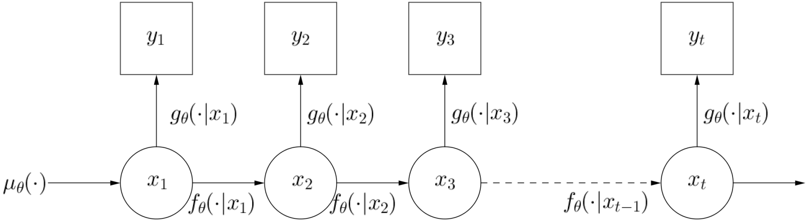
(b) Run an SMC algorithm targeting $p_{\theta^*}(x_{1:T}|y_{1:T})$

(c) Sample $X^*_{1:T} \sim \hat{p}_{\theta^*}(\cdot|y_{1:T})$ and let $\hat{p}_{\theta^*}(y_{1:T})^*$ denote the corresponding marginal likelihood estimate.

(d) Accept update with probability

$$1 \wedge \frac{\hat{p}_{\theta^*}(y_{1:T})p(\theta^*)}{\hat{p}_{\theta(i-1)}(y_{1:T})p\{\theta(i-1)\}} \frac{q\{\theta(i-1)|\theta^*\}}{q\{\theta^*|\theta(i-1)\}} \tag{5}$$

Application



Data-generating model

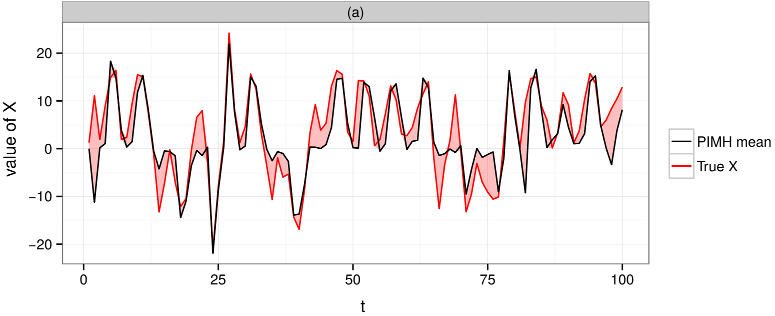
μ: $X_1 \sim \mathcal{N}(0, 5)$

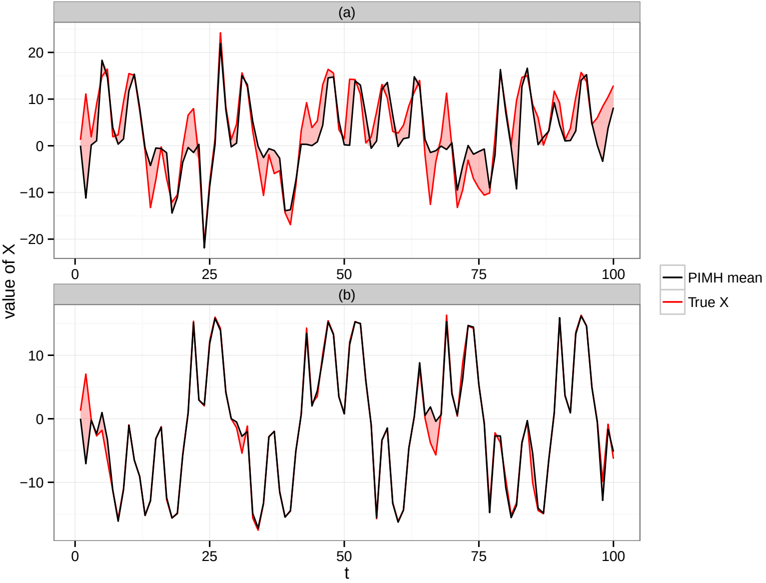
f_θ: $X_t = \frac{X_{t-1}}{2} + 25\frac{X_{t-1}}{1+X_{t-1}^2} + 8\cos(1.2t) + V_t$

g_θ: $Y_t = \frac{X_t^2}{20} + W_t$

with $V_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_V^2)$ and $W_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_W^2)$.

$\theta =$ (σ_V, σ_W)


(a)
value of X
20
10
0
−10
−20
0
25
50
75
100
t
PIMH mean
True X


(a)
(b)
value of X
t
PIMH mean
True X
20
10
0
−10
−20
0
25
50
75
100

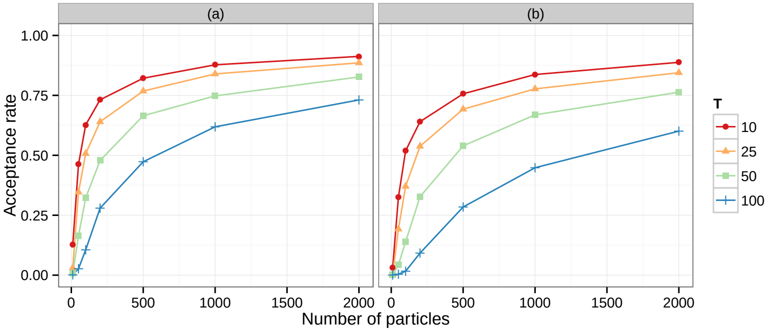


Figure: Acceptance rates from PIMH sampler different values of T and θ, (a) $\sigma_V^2 = \sigma_W^2 = 10$, (b) $\sigma_V^2 = 10$ and $\sigma_W^2 = 1$

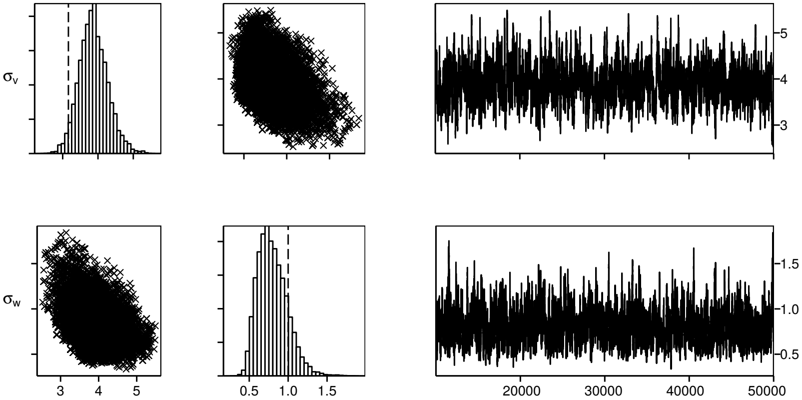


Figure: Posterior distribution for σ_V and σ_W from PMMH sampler

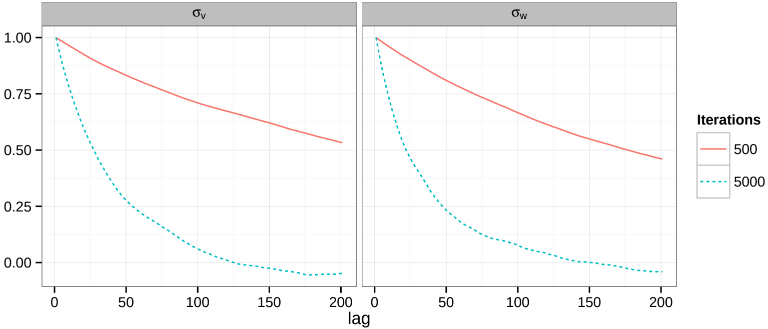


Figure: Autocorrelation function for N = 500 and N = 5000 particles