UPPSALA
UNIVERSITET

Particle Metropolis-Hastings

Johan Dahlin
uu@johandahlin.com

Department of Information Technology,
Uppsala University, Sweden.

February 16, 2017

UPPSALA UNIVERSITET

Short Bio

2005-2011: MSc in Engineering Physics, Umeå University.

2009-2011: BSc in Economics, Umeå University.

2011-2016: PhD in Automatic Control, Linköping University

2014: Visiting researcher at UNSW Business School, Sydney.

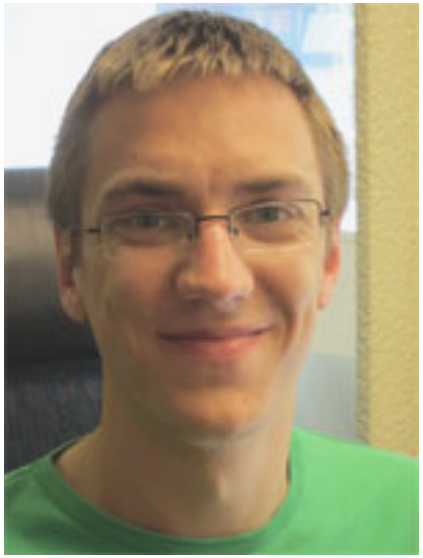
2016-2017: PostDoc at Sectra AB/Uppsala University.

March 2017: PostDoc at STIMA, Linköping University.

UPPSALA
UNIVERSITET

This is collaborative work together with

Dr. Fredrik Lindsten (Uppsala University, Sweden).
Prof. Thomas Schön (Uppsala University, Sweden).



What are we going to do?

- Give a (hopefully) gentle introduction to (P)MCMC.
- Develop some intuition for PMH and its pros/cons.
- Discuss some recent developments.

Why are we doing this?

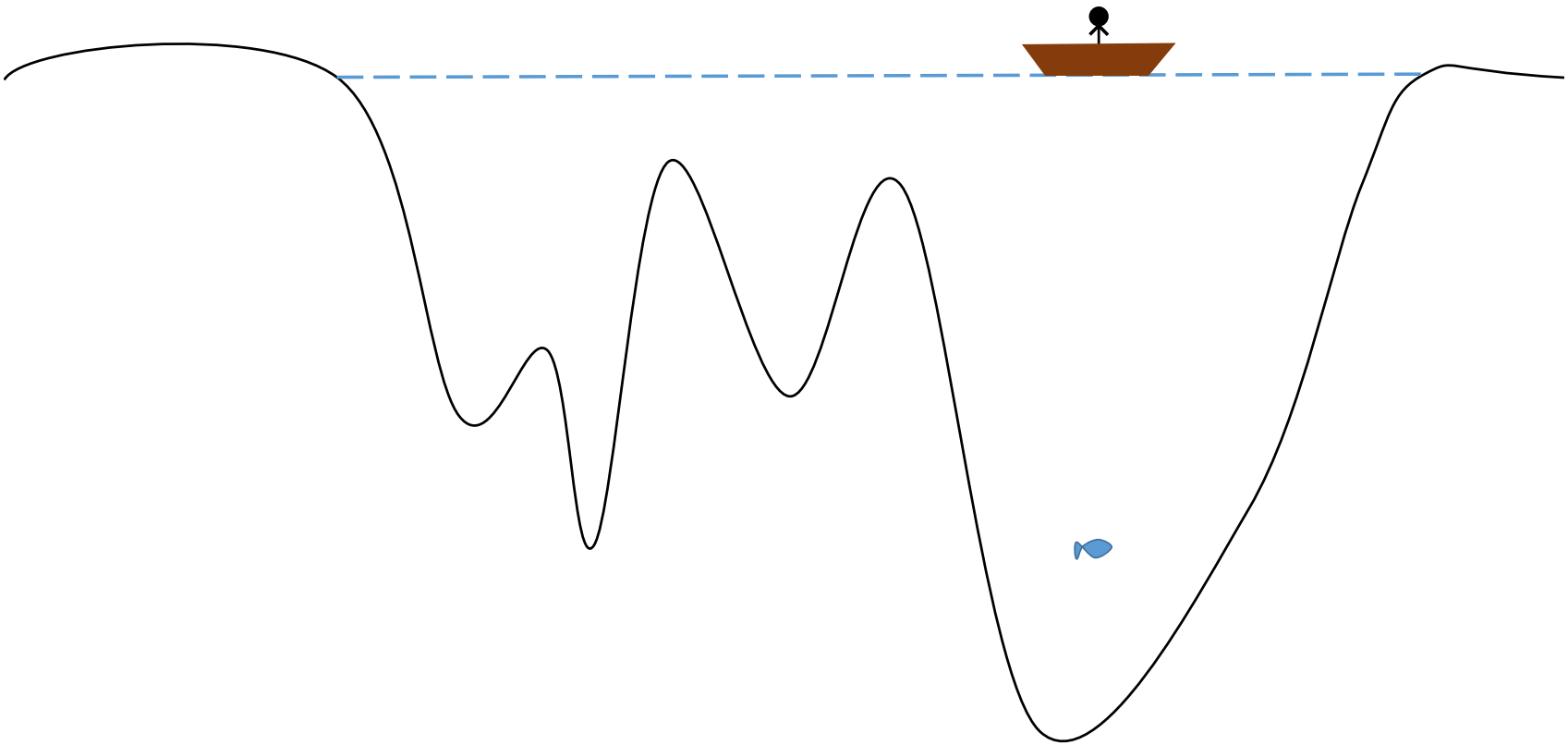
- PMH is general algorithm for Bayesian inference in SSMs.
- Relatively simple to implement and tune.

How will we do this?

- Employ intuition and analogues with optimisation.
- Investigate PMH using simulations and not maths.
- Illustrate PMH on a real-world example.
- By asking questions.

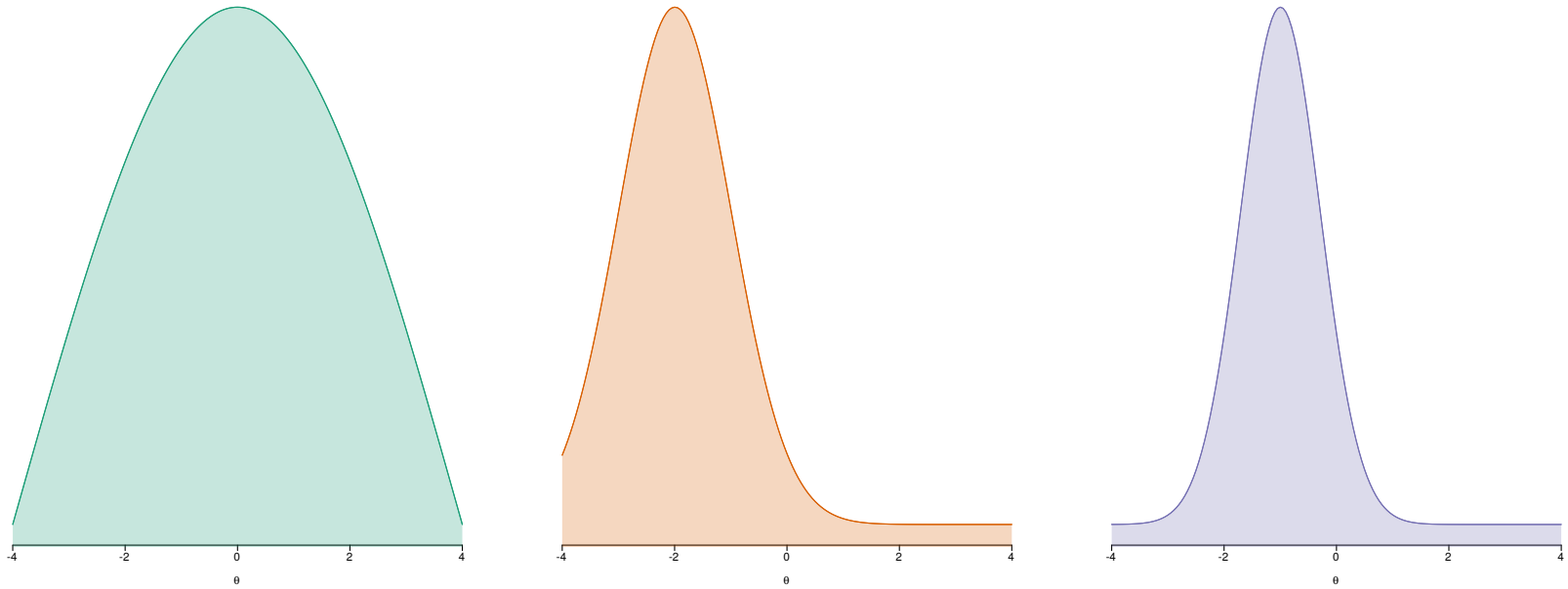
UPPSALA UNIVERSITET

Mapping a lake



UPPSALA UNIVERSITET

Bayesian parameter inference



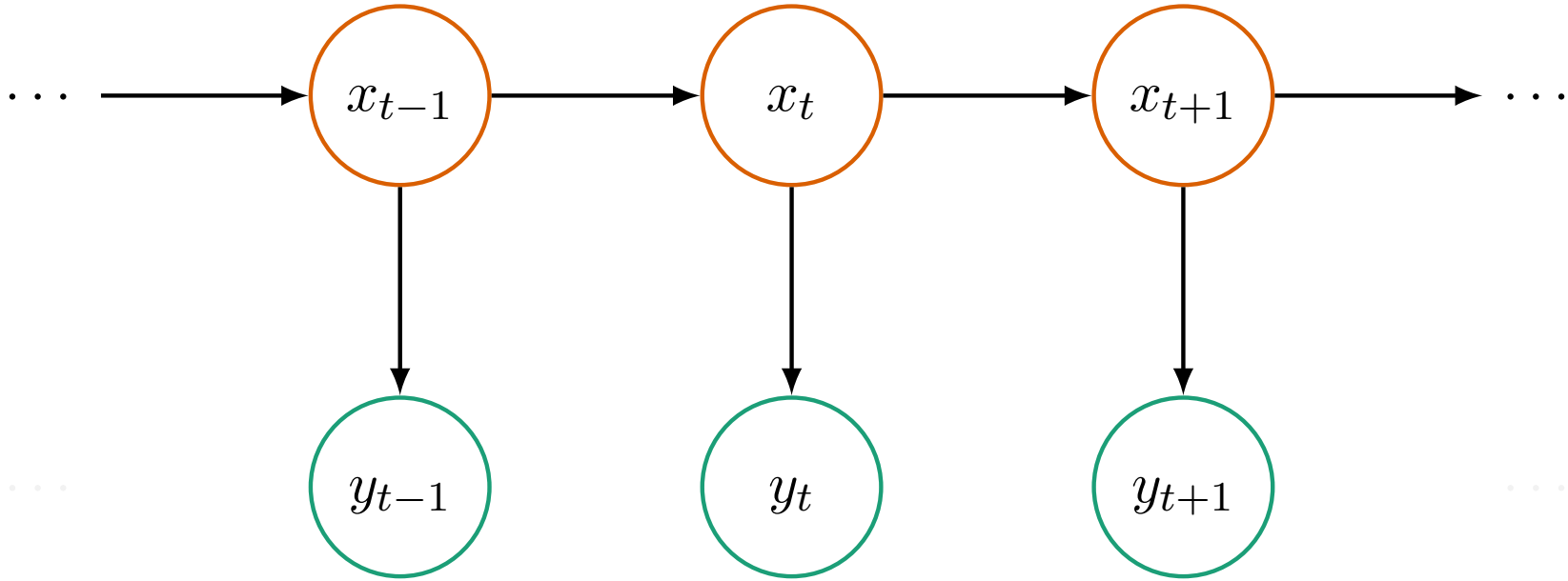
$$\pi(\theta) = p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)},$$

$$\pi[\varphi] = \mathbb{E}_\pi\left[\varphi(\theta)\right] = \int \varphi(\theta')\pi(\theta')\,\mathrm{d}\theta'.$$

UPPSALA UNIVERSITET

State space models [I/II]

Markov chain $[X_{0:T}, Y_{1:T}]$ with $X_t \in \mathcal{X} = \mathbb{R}$, $Y_t \in \mathcal{Y} = \mathbb{R}$, $t \in \mathbb{N}$.



$x_0 \sim \mu_\theta(x_0)$ $\quad$ $x_{t+1}|x_t \sim f_\theta(x_{t+1}|x_t)$, $\quad$ $y_t|x_t \sim g_\theta(y_t|x_t)$.

UPPSALA UNIVERSITET

State space models [II/II]

Linear Gaussian SSM with $\theta = [\mu, \phi, \sigma_v, \sigma_e]$ is given by

$$x_{t+1}|x_t \sim \mathcal{N}\left(x_{t+1}; \mu + \phi(x_t - \mu), \sigma_v^2\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; x_t, \sigma_e^2\right).$$

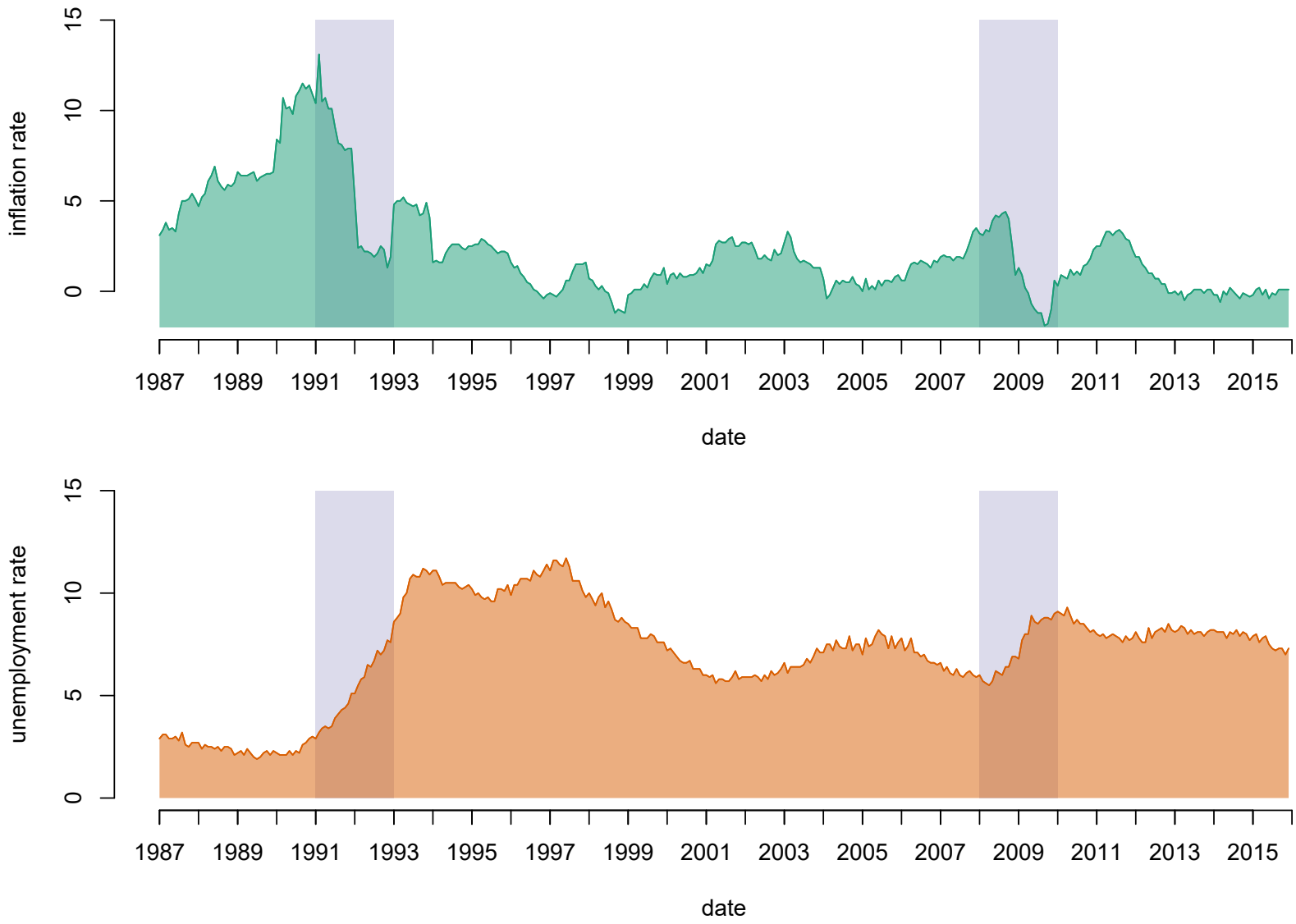
or equivalently by

$$x_{t+1} = \mu + \phi(x_t - \mu) + \sigma_v v_t,$$
$$y_t = x_t + \sigma_e e_t,$$

with $v_t, e_t \overset{iid}{\sim} \mathcal{N}(0, 1)$.

UPPSALA UNIVERSITET

Philips curve: Swedish data



UPPSALA UNIVERSITET

Philips curve: state space model

Inflation rate y_t and unemployment rate u_t

$$x_t|x_{t-1} \sim \mathcal{N}\left(x_t; \phi x_{t-1} + \mu(x_{t-1}), \sigma_v^2(x_{t-1}, u_{t-1})\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; y_{t-1} + \beta(u_t - x_t), \sigma_e^2\right),$$

where $\theta = \{\phi, \alpha, \beta, \sigma_e\}$ and

$$\mu(x_{t-1}) = \alpha\Big[1 + \exp(-x_{t-1})\Big]^{-1},$$
$$\sigma_v^{-1}(x_{t-1}, u_{t-1}) = 1 + \exp\Big(-|u_{t-1} - x_{t-1}|\Big),$$

x_t denotes the NAIRU (structural unemployment rate).

Exploring posteriors by Markov chains

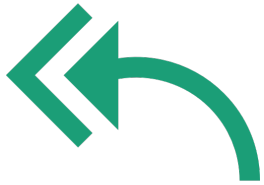
UPPSALA UNIVERSITET

Markov chains: basic properties

A sequence of random variables $\{X_k\}_{k=0}^{K}$ with the property

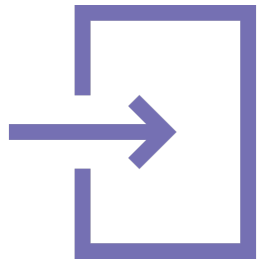
$$\mathbb{P}[X_k \in A | x_{0:k-1}] = \mathbb{P}[X_k \in A | x_{k-1}] = \int_A R(x_{k-1}, x_k)\, \mathrm{d}x_k.$$

We will consider ergodic chains with the properties:



Reach any point (irreducible)

No cycles (aperiodic)

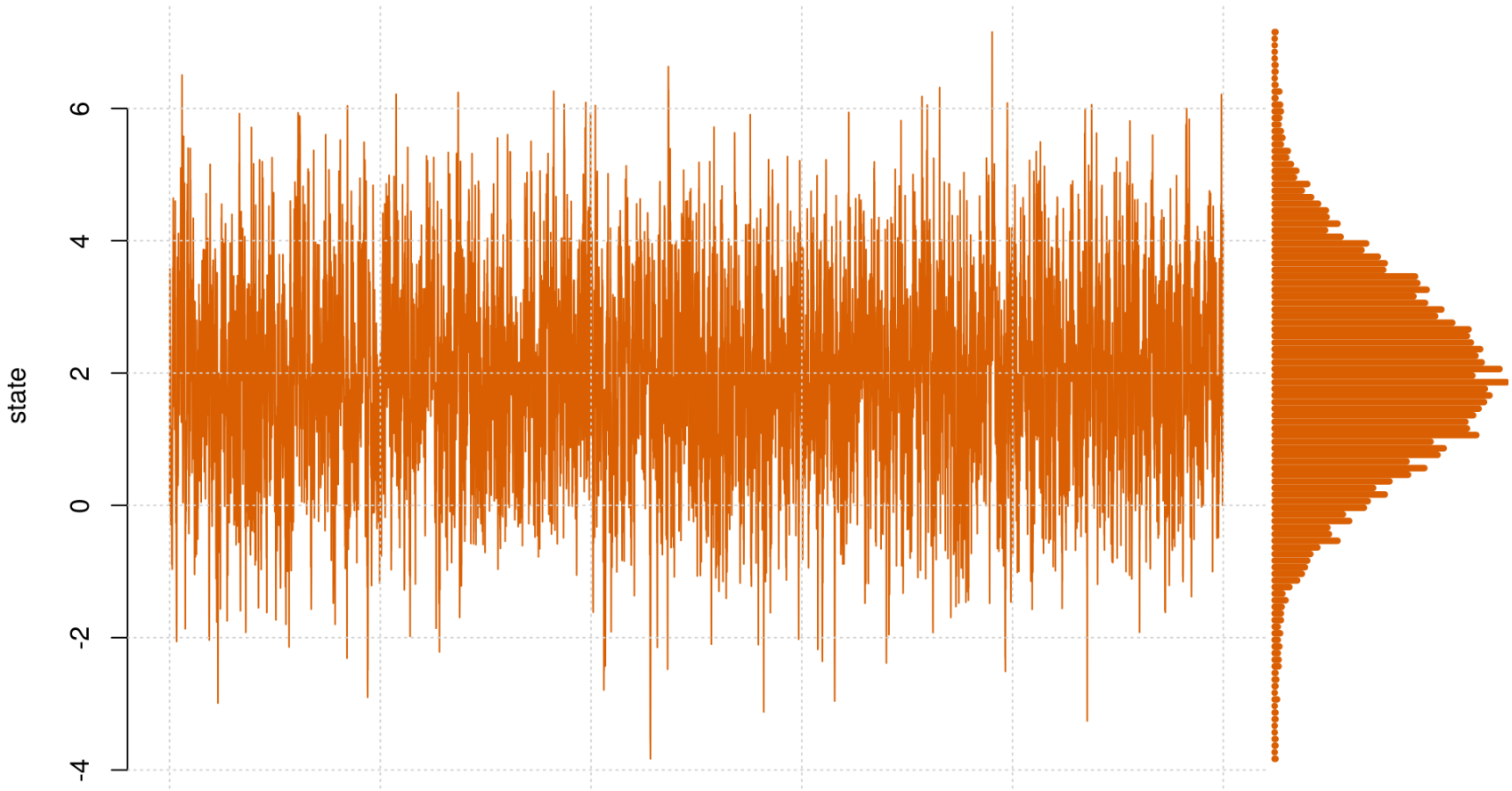


Does not get stuck (recurrent)

Icons made by Freepik from `www.flaticon.com`.

UPPSALA UNIVERSITET

Markov chains: stationary distribution



$$\theta_k|\theta_{k-1} \sim \mathcal{N}\left(\theta_k; \mu + \phi(\theta_{k-1} - \mu), \sigma^2\right).$$

UPPSALA UNIVERSITET

Metropolis-Hastings: algorithm

Initialise in θ_0 and then generate samples $\{\theta_k\}_{k=1}^K$ from $\pi(\theta)$ by

(i) Sample a candidate parameter θ' by

$$\theta' \sim \mathcal{N}(\theta'; \theta_{k-1}, \Sigma).$$

(ii) Accept θ' by setting $\theta_k \leftarrow \theta'$ with probability

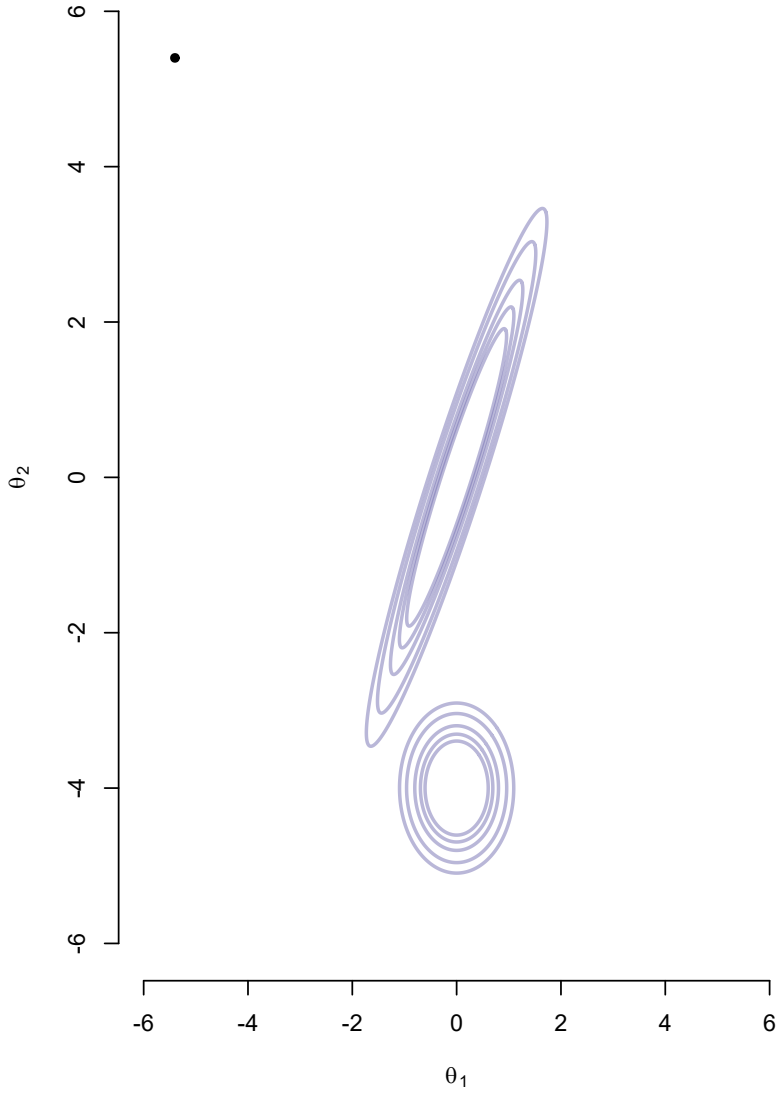
$$\min\left\{1, \frac{\pi(\theta')}{\pi(\theta_{k-1})}\right\} = \min\left\{1, \frac{p(\theta')}{p(\theta_{k-1})}\frac{p(y|\theta')}{p(y|\theta_{k-1})}\frac{p(y)}{p(y)}\right\}$$

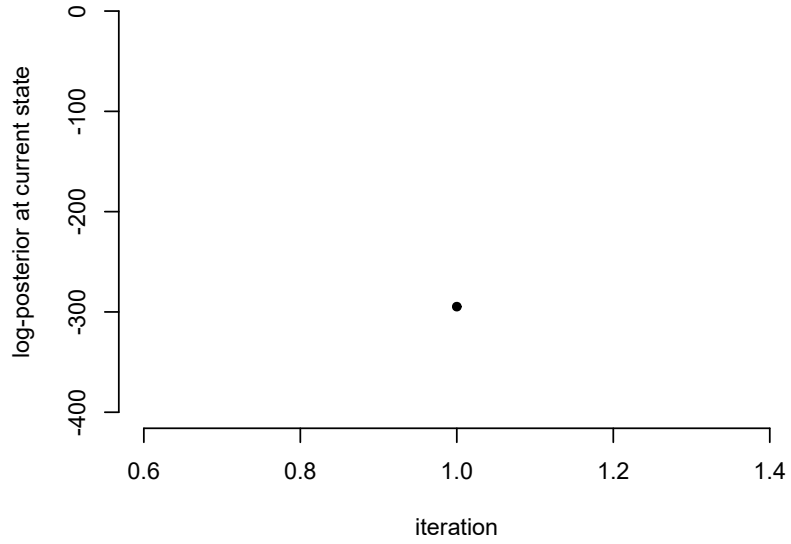
and otherwise reject θ' by setting $\theta_k \leftarrow \theta_{k-1}$.

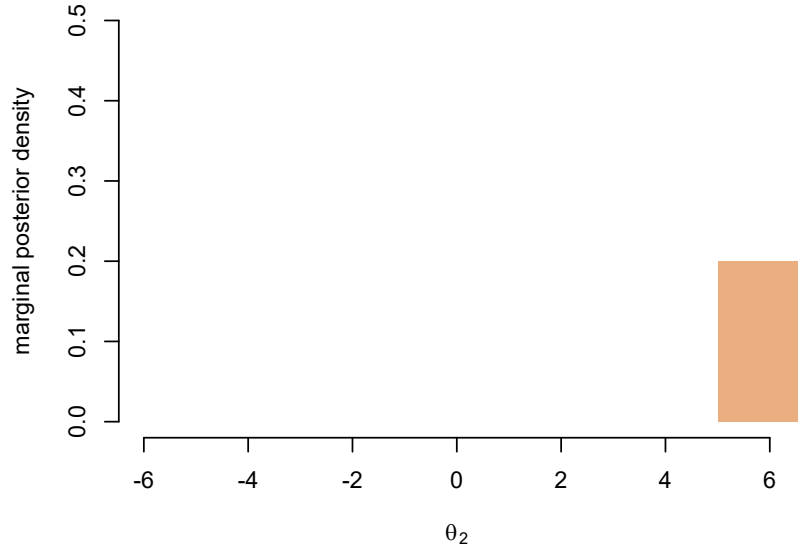
User choices: K and Σ.

UPPSALA UNIVERSITET

Metropolis-Hastings: toy example

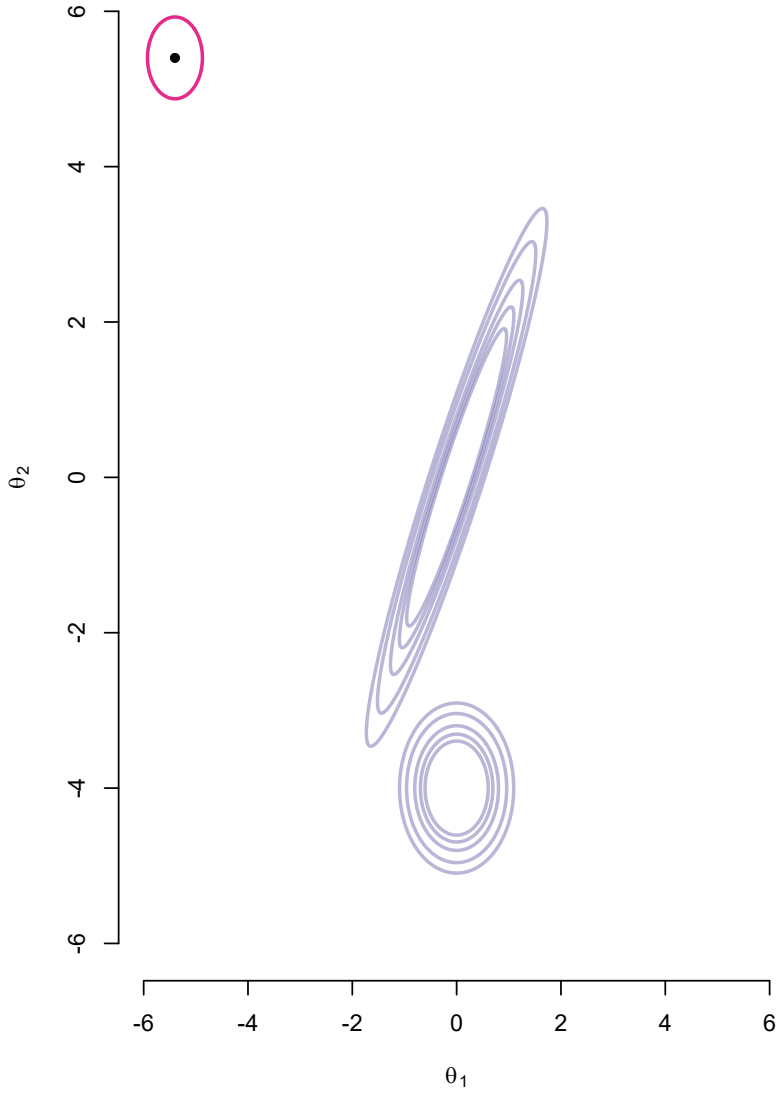


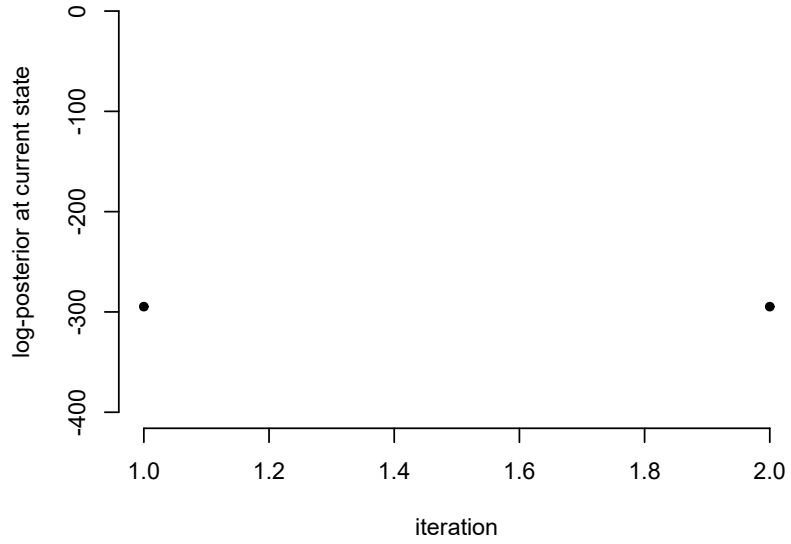


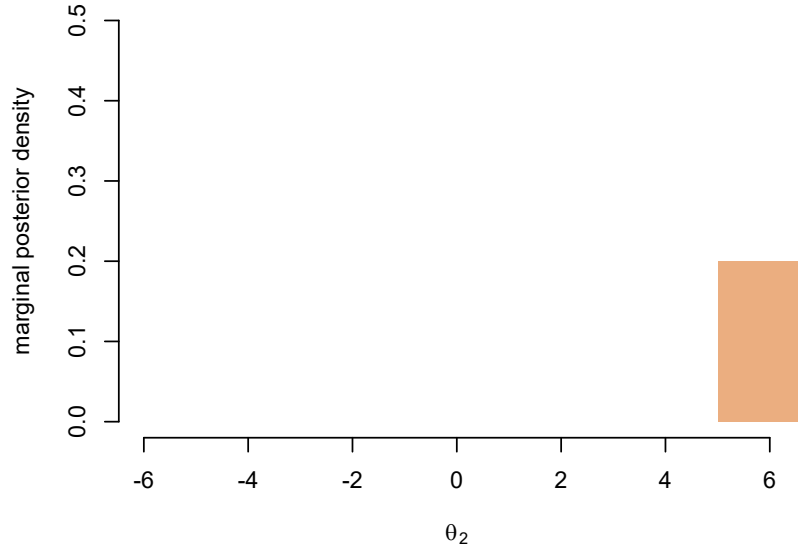


UPPSALA UNIVERSITET

Metropolis-Hastings: toy example

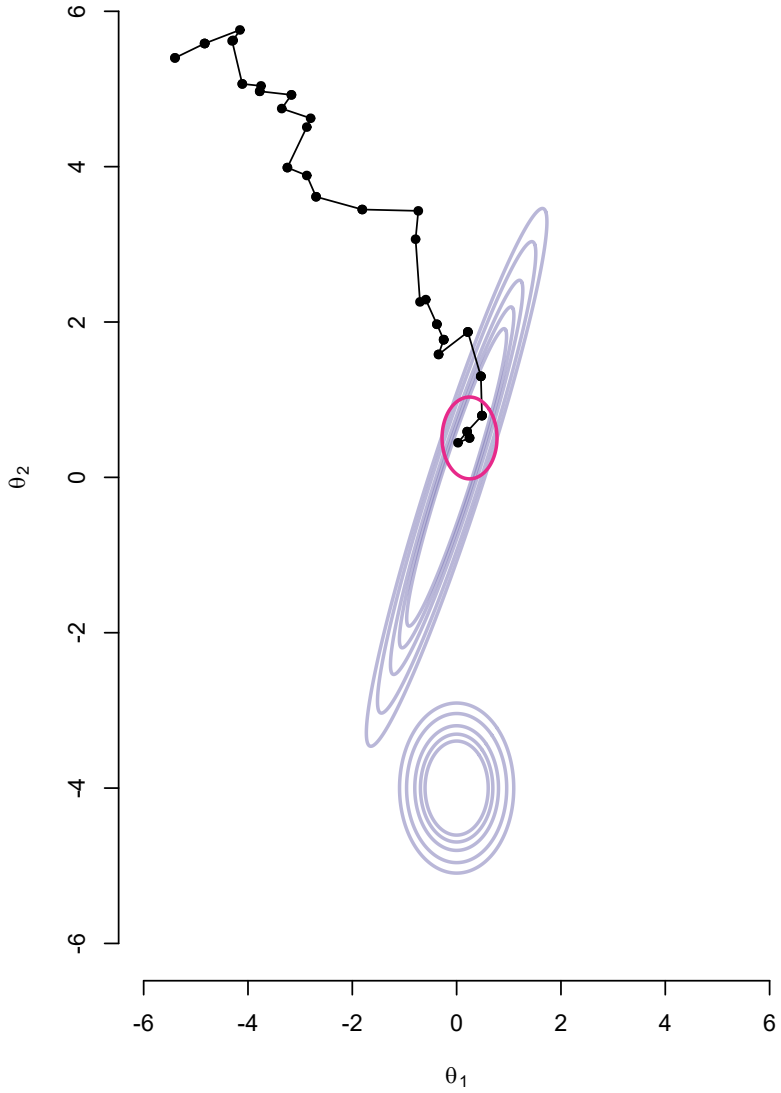


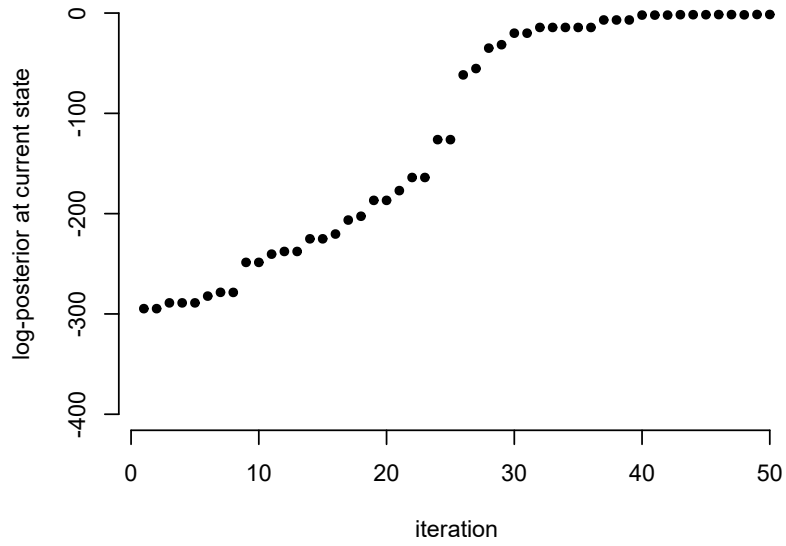


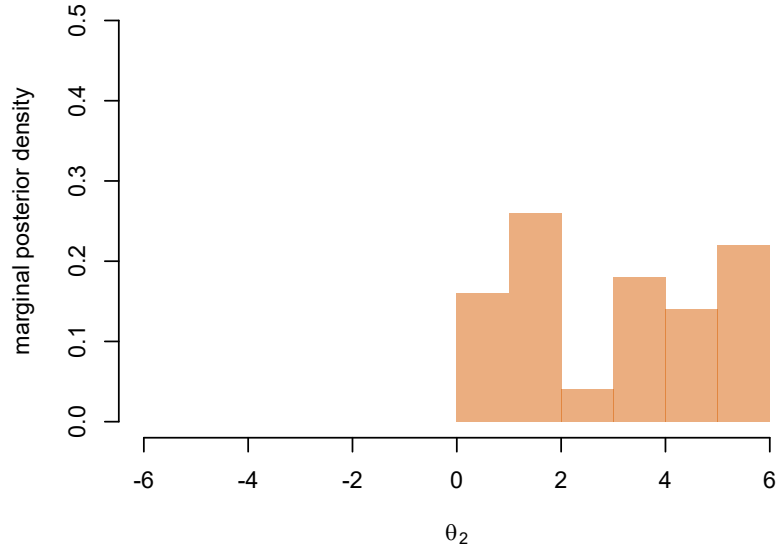


UPPSALA UNIVERSITET

Metropolis-Hastings: toy example

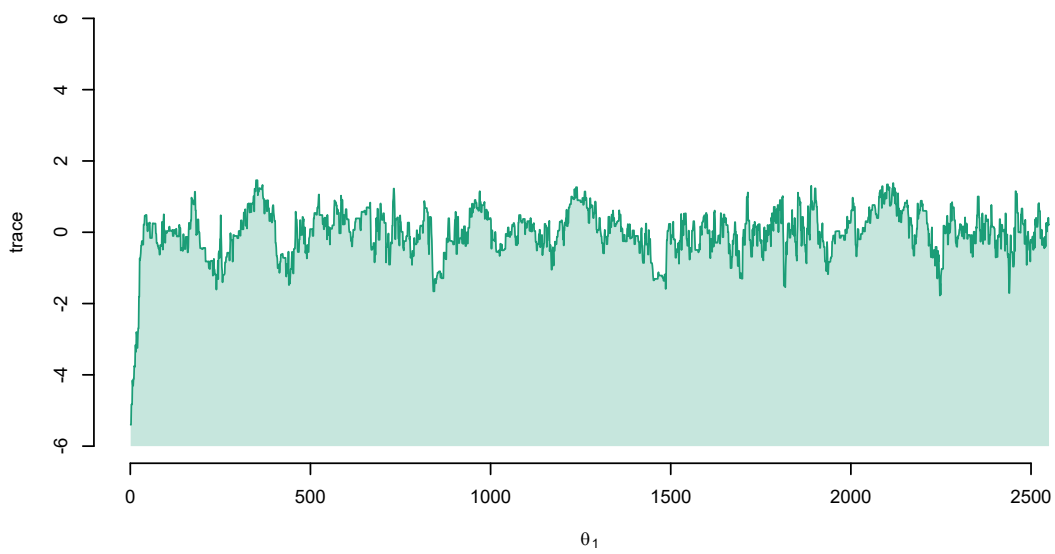


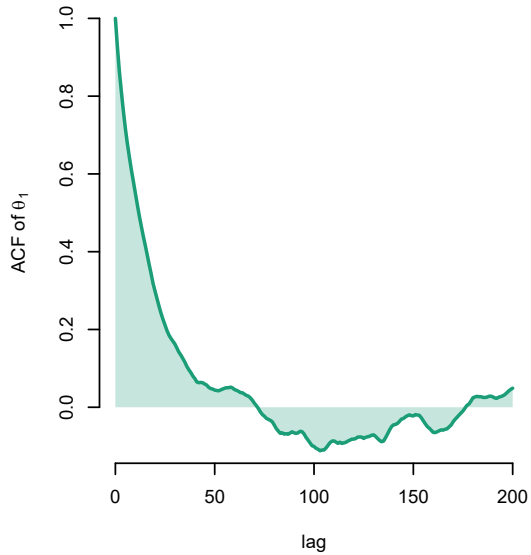


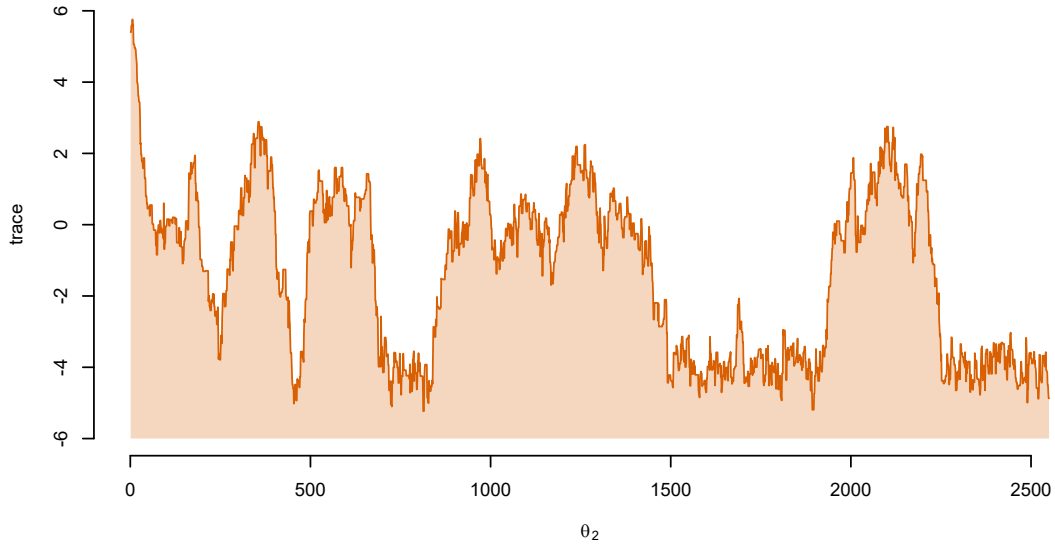


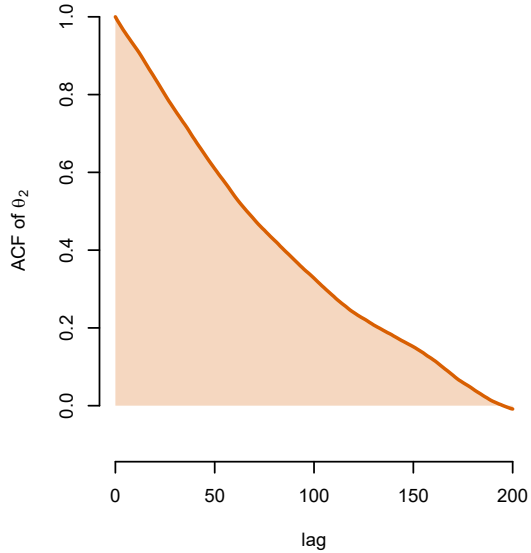
UPPSALA UNIVERSITET

Metropolis-Hastings: proposal and mixing



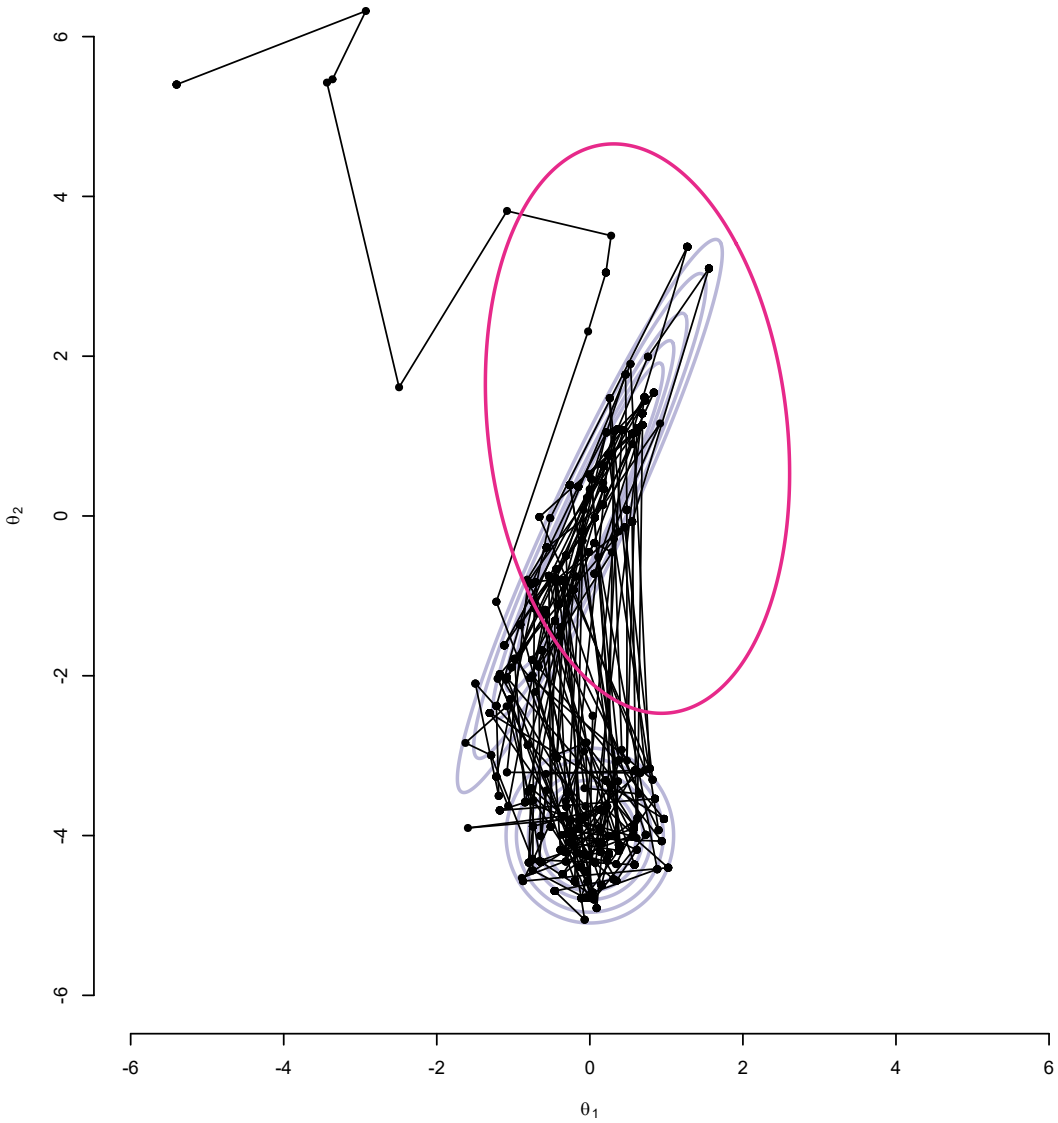


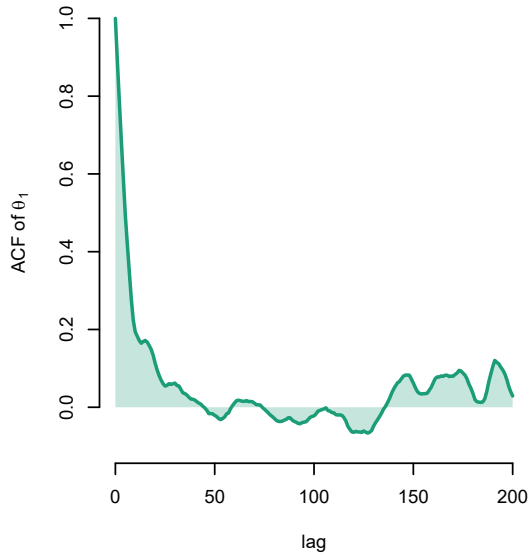


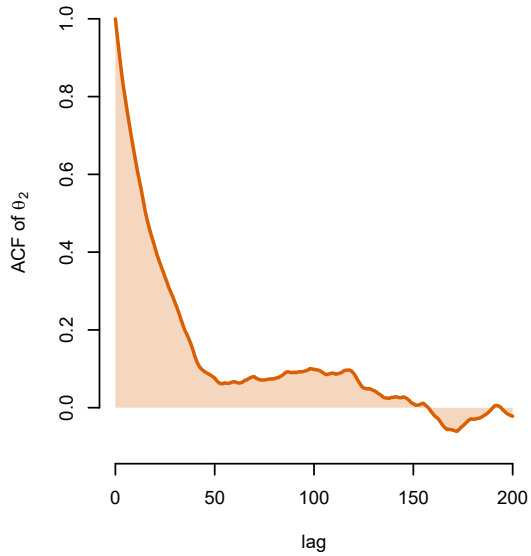


UPPSALA UNIVERSITET

Metropolis-Hastings: proposal and mixing







UPPSALA UNIVERSITET

Metropolis-Hastings: theoretical results

The expectation of a test function φ wrt. $\pi(\theta) = p(\theta|y)$,

$$\pi[\varphi] = \mathbb{E}_\pi\left[\varphi(\theta)\right] = \int \varphi(\theta')\pi(\theta')\,\mathrm{d}\theta',$$

can be approximated using MCMC by

$$\widehat{\pi}^K[\varphi] = \frac{1}{K}\sum_{k=1}^{K}\varphi(\theta_k),$$

which under geometric ergodicity obeys

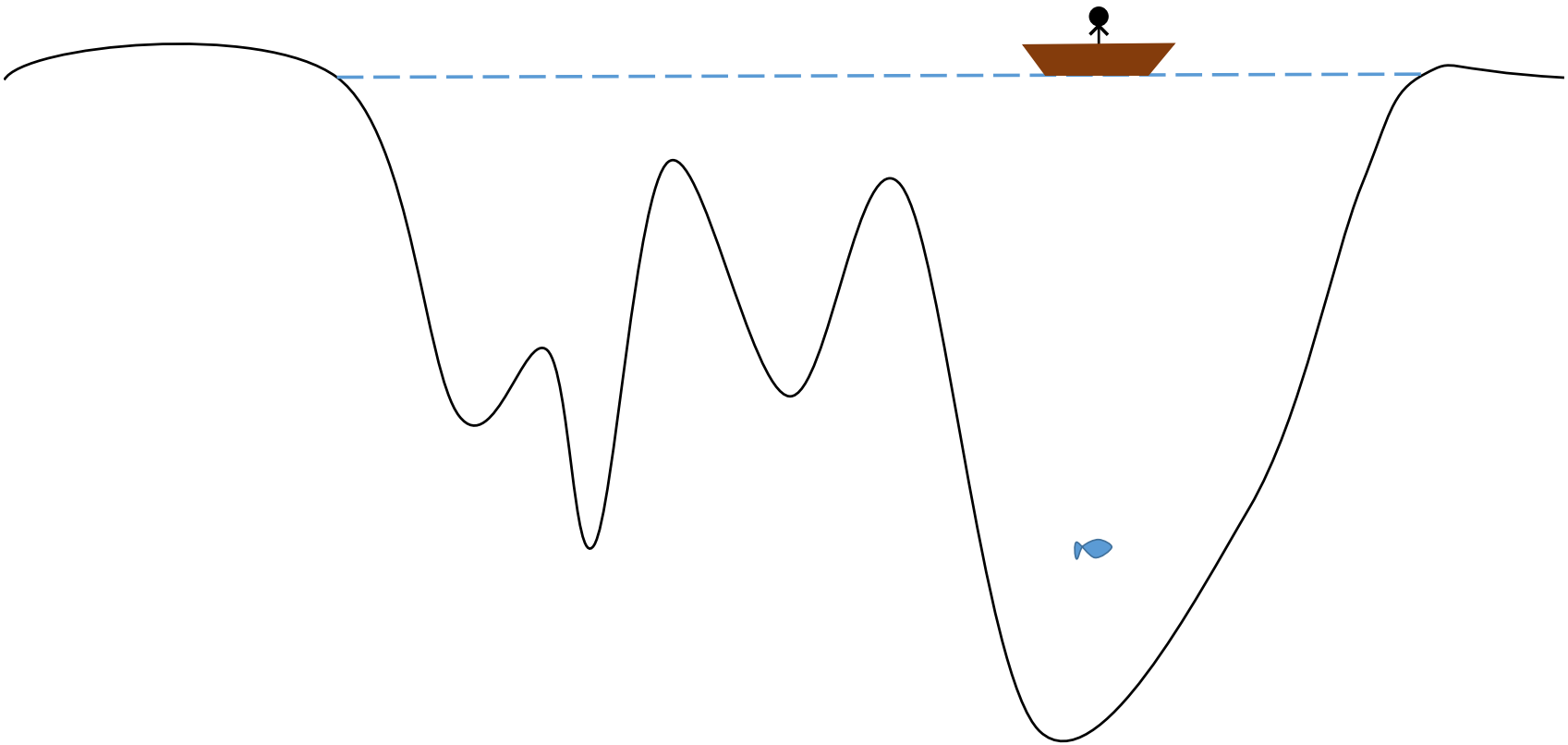
$$\sqrt{K}\left[\pi[\varphi] - \widehat{\pi}^K[\varphi]\right] \xrightarrow{d} \mathcal{N}\left(0, \sigma_\varphi^2 \cdot \mathsf{IACT}(\theta_{1:K})\right),$$

when $K \to \infty$.

Approximating the target by particles

UPPSALA
UNIVERSITET

Mapping a stormy lake



UPPSALA UNIVERSITET

Why do we need a particle filter?

For a state-space model,

$$x_0 \sim \mu_\theta(x_0), \quad x_t|x_{t-1} \sim f_\theta(x_t|x_{t-1}), \quad y_t|x_t \sim g_\theta(y_t|x_t),$$

the likelihood is given by

$$p(y|\theta) = p(y_1|\theta) \prod_{t=2}^{T} \int g_\theta(y_t|x_t) f_\theta(x_t|x_{t-1}) p_\theta(x_{t-1}|y_{1:t-1}) \, \mathrm{d}x_{t:t-1},$$

where the particle filter can estimate

$$\widehat{p}_\theta(x_{t-1}|y_{1:t-1}) = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{x_{t-1}^{(i)}}(x_{t-1}).$$

UPPSALA UNIVERSITET

Particle Metropolis-Hastings (PMH)

Initialise in θ_0 and generate samples $\{\theta_k\}_{k=1}^K$ from $\pi(\theta)$ by

(i) Sample candidate parameter $\theta' \sim \mathcal{N}(\theta'; \theta_{k-1}, \Sigma)$.

(ii) Estimate posterior $\widehat{\pi}^N(\theta') = \texttt{ParticleFilter}(\theta', N)$.

(iii) Accept θ' by setting $\theta_k \leftarrow \theta'$ with probability

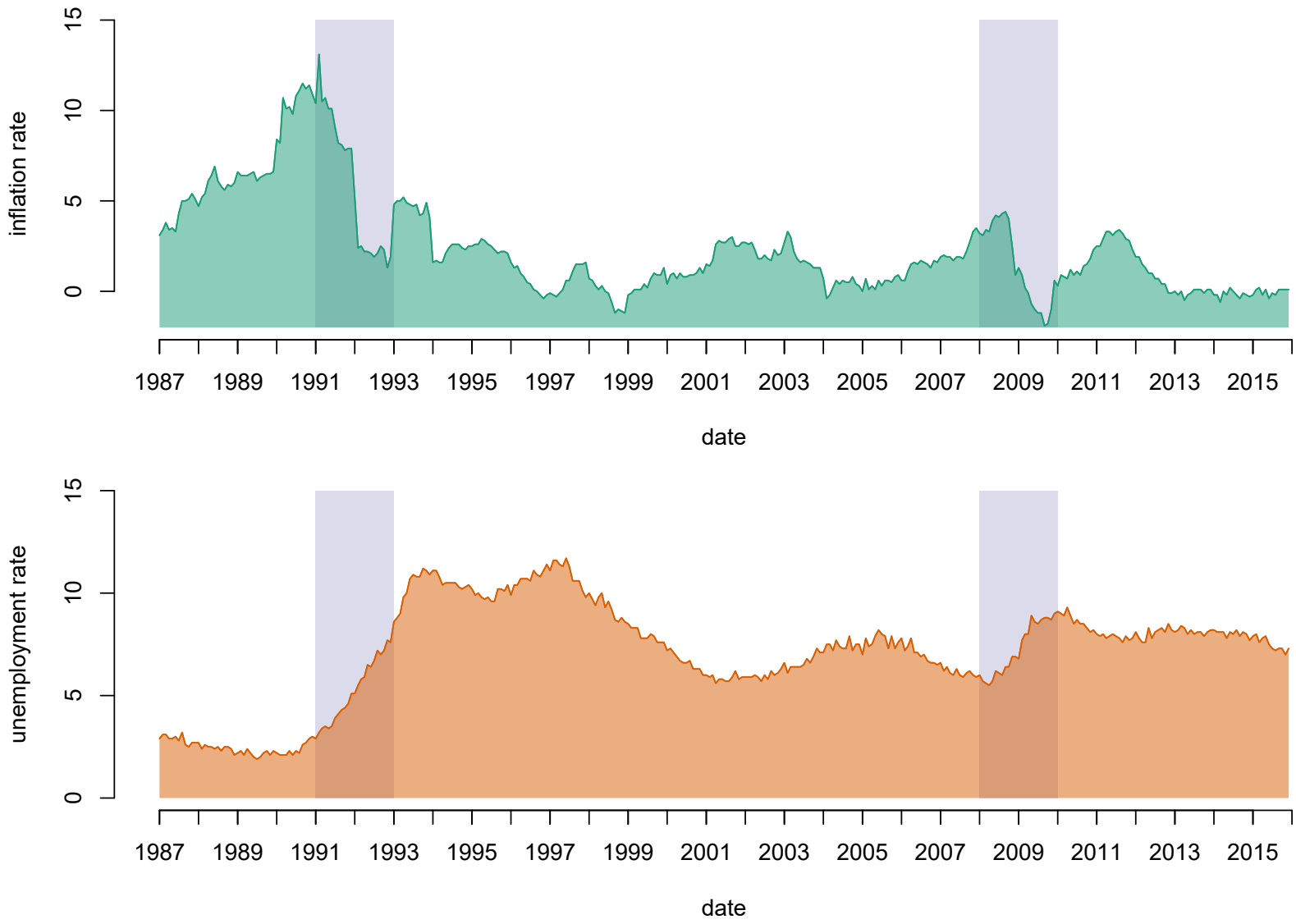
$$\min\left\{1, \frac{\widehat{\pi}^N(\theta')}{\widehat{\pi}^N(\theta_{k-1})}\right\},$$

and otherwise reject θ' by setting $\theta_k \leftarrow \theta_{k-1}$.

User choices: K, Σ and N.

UPPSALA UNIVERSITET

Philips curve: Swedish data



UPPSALA UNIVERSITET

Philips curve: state space model

Inflation rate y_t and unemployment rate u_t

$$x_t|x_{t-1} \sim \mathcal{N}\left(x_t; \phi x_{t-1} + \mu(x_{t-1}), \sigma_v^2(x_{t-1}, u_{t-1})\right),$$
$$y_t|x_t \sim \mathcal{N}\left(y_t; y_{t-1} + \beta(u_t - x_t), \sigma_e^2\right),$$

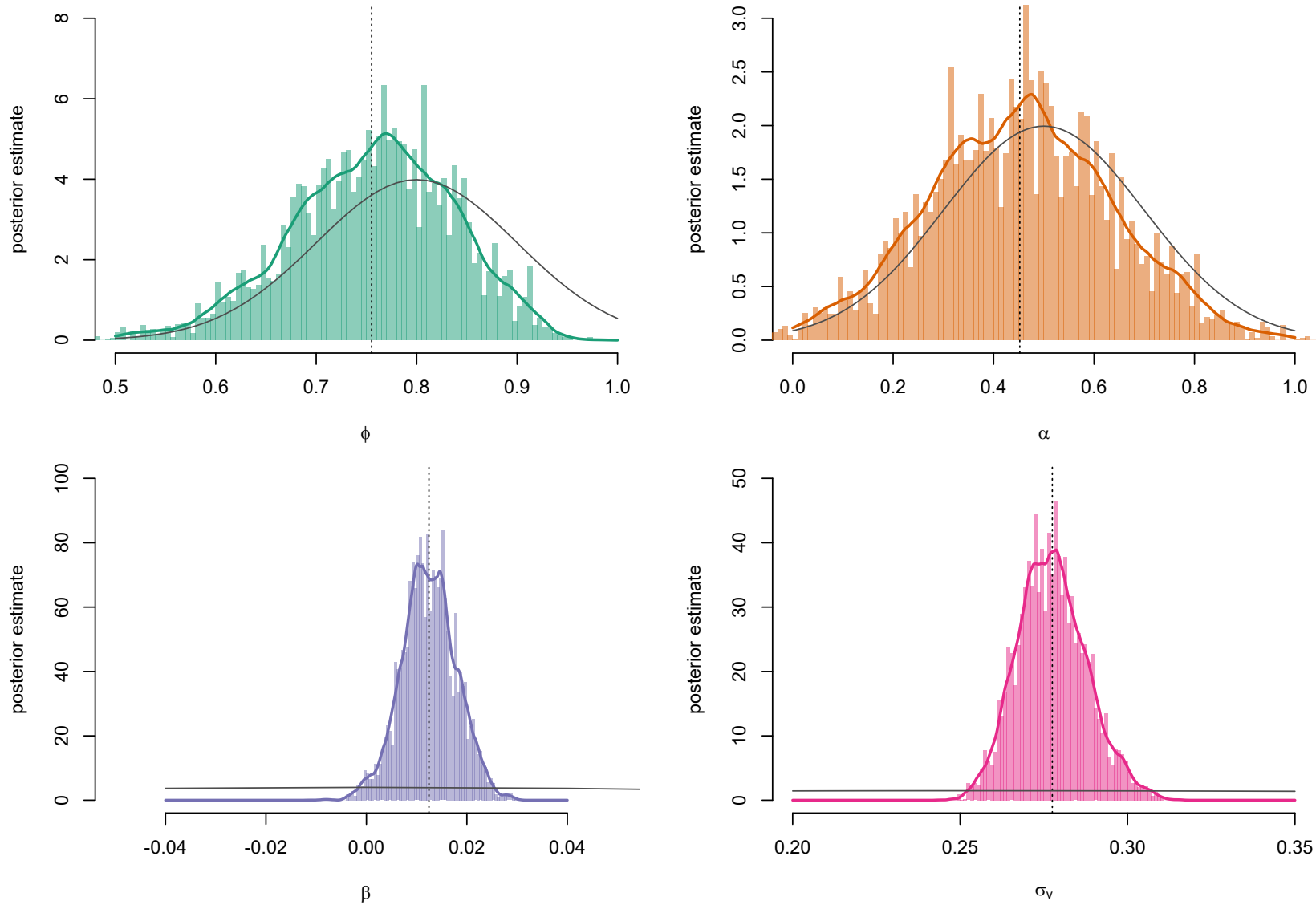
where $\theta = \{\phi, \alpha, \beta, \sigma_e\}$ and

$$\mu(x_{t-1}) = \alpha\Big[1 + \exp(-x_{t-1})\Big]^{-1},$$
$$\sigma_v^{-1}(x_{t-1}, u_{t-1}) = 1 + \exp\Big(-|u_{t-1} - x_{t-1}|\Big),$$

x_t denotes the NAIRU (structural unemployment rate).

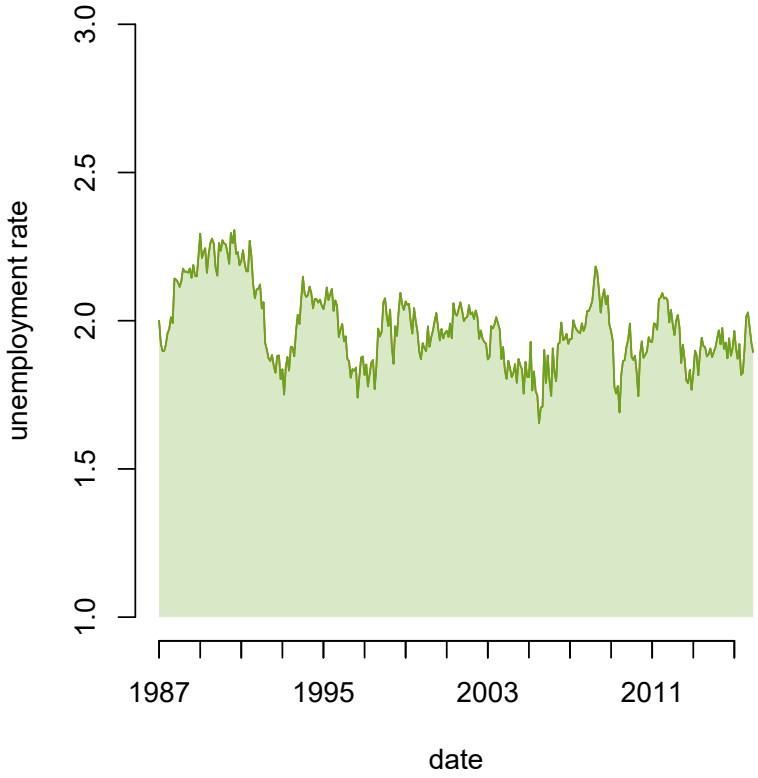
UPPSALA UNIVERSITET

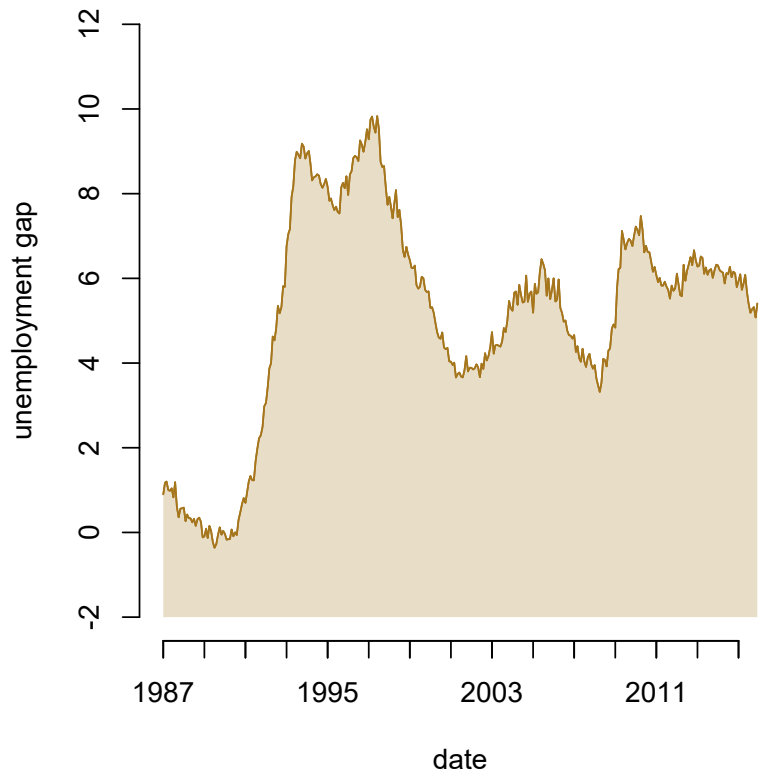
Philips curve: posterior estimates



UPPSALA UNIVERSITET

Philips curve: state estimates

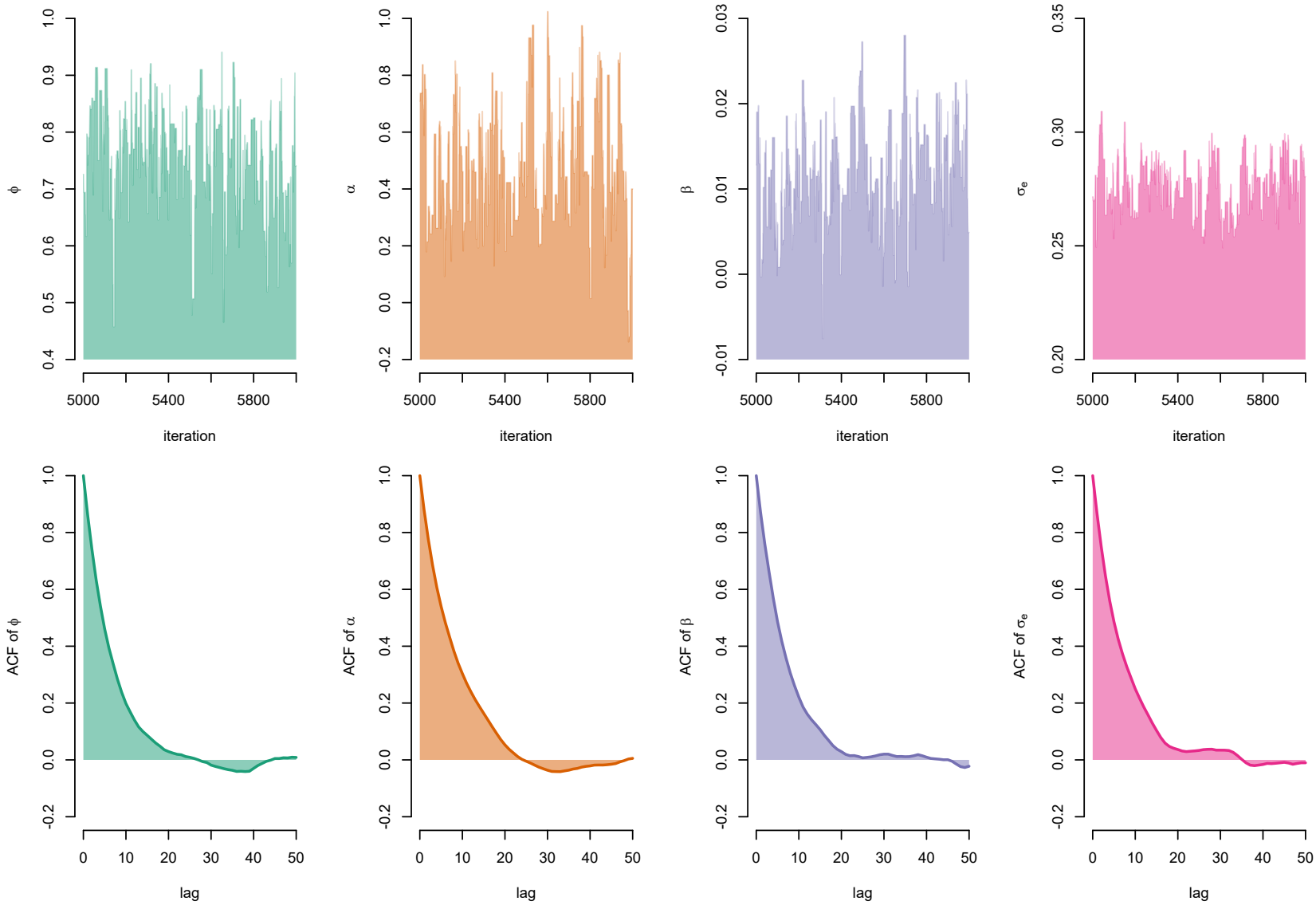




Improving the PMH algorithm

UPPSALA UNIVERSITET

Philips curve: trace plots



UPPSALA UNIVERSITET

Correlating particle filters [I/II]

Introduce the random variables used by the particle filter

$$\widehat{\pi}^N(\theta') = \texttt{ParticleFilter}(\theta', N, \mathbf{z}),$$

into the state vector of the Markov chain, i.e.

$$\boldsymbol{\theta} = \{\theta, \mathbf{z}\},$$

with the (symmetric) Crank-Nicolson proposal

$$q(\mathbf{z}'|\mathbf{z}) = \mathcal{N}\left(\mathbf{z}'; \sqrt{1-\sigma_z^2}\mathbf{z}, \sigma_z^2 \mathbf{I}_{|\mathbf{z}|}\right),$$

for some step size σ_z.

UPPSALA UNIVERSITET

Correlating particle filters [II/II]

Gives corr$(\widehat{\pi}^N(\theta'), \widehat{\pi}^N(\theta_{k-1})) > 0$ and variance reduction in

$$\min\left\{1, \frac{\widehat{\pi}^N(\theta')}{\widehat{\pi}^N(\theta_{k-1})}\right\} = \min\left\{1, \frac{\pi(\theta') + \epsilon'}{\pi(\theta_{k-1}) + \epsilon_{k-1}}\right\},$$

where corr$(\epsilon', \epsilon_{k-1}) > 0$.

Different scaling: $N \propto T^{\eta}$ with $\eta < 1$.

UPPSALA UNIVERSITET

Better informed proposals [I/III]

Introduce mode-seeking drift in the proposal by the gradient

$$\mathcal{G}(\theta') = \nabla \log \pi(\theta)\big|_{\theta=\theta'},$$

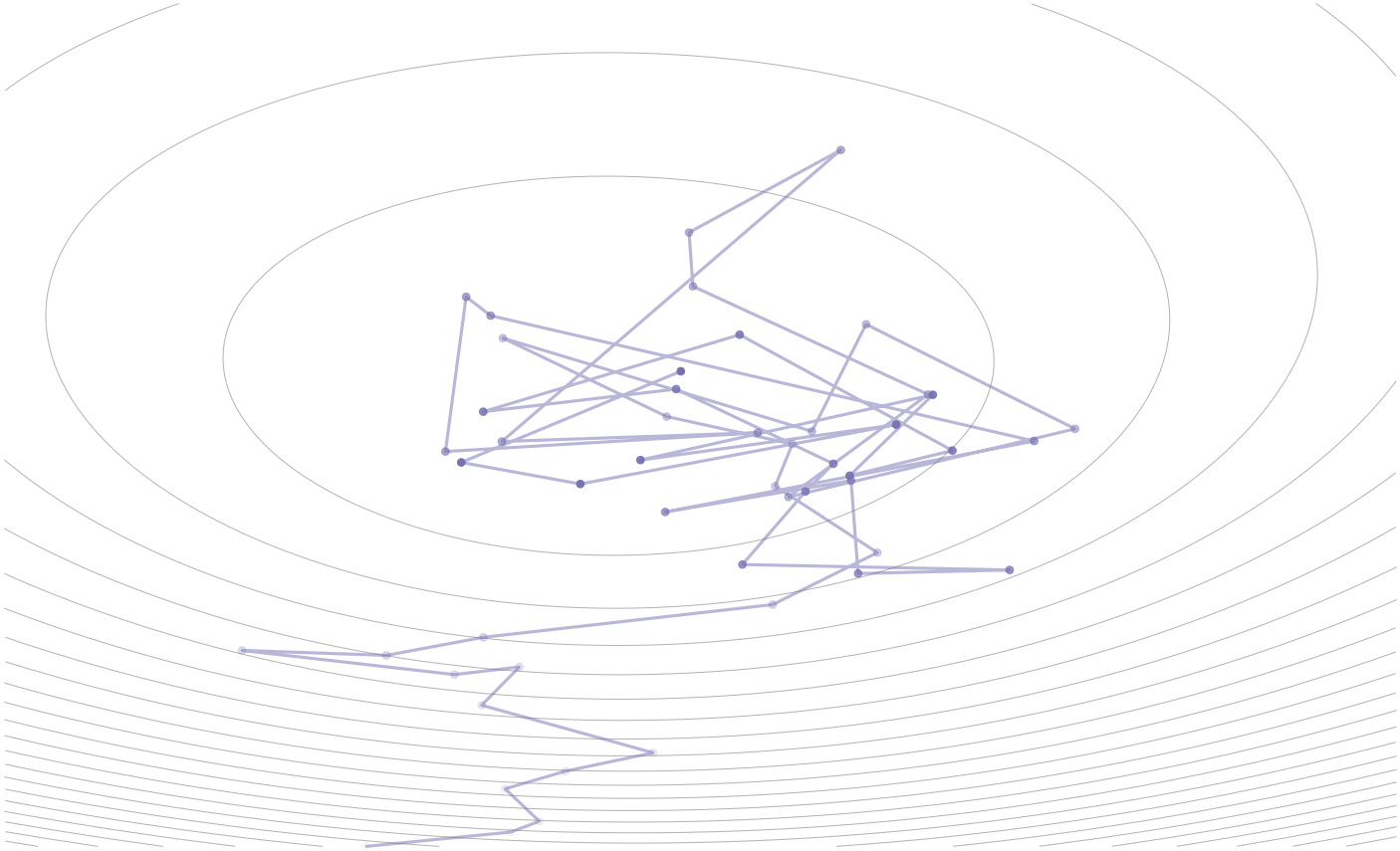
scaled by the negative Hessian (observed information matrix)

$$\mathcal{H}(\theta') = -\nabla^2 \log \pi(\theta)\big|_{\theta=\theta'},$$

which acts like a position-dependent step-size.

UPPSALA UNIVERSITET

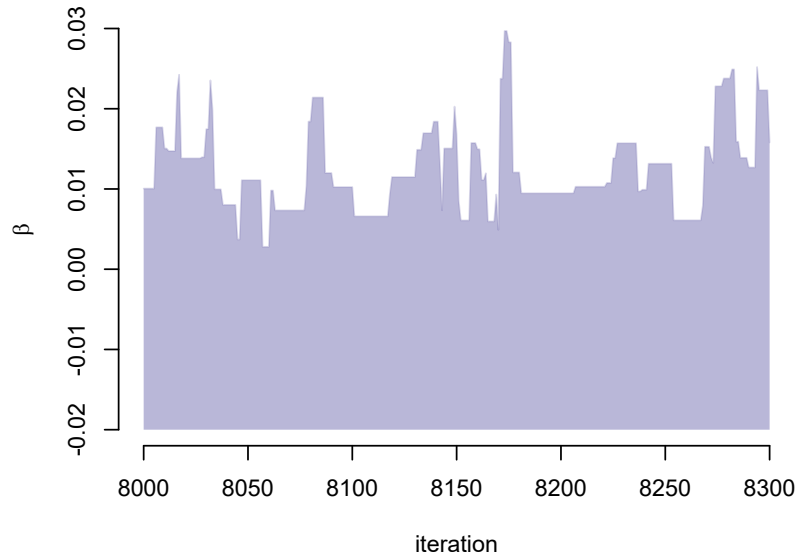
Better informed proposals [II/III]

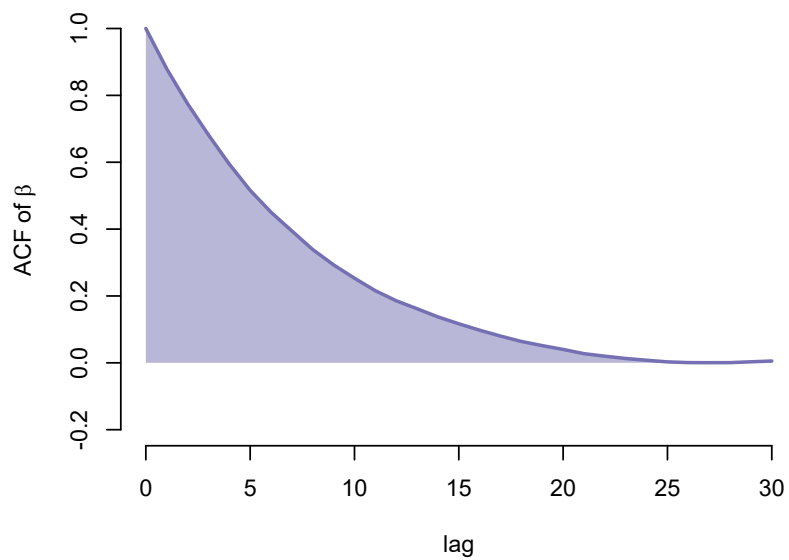


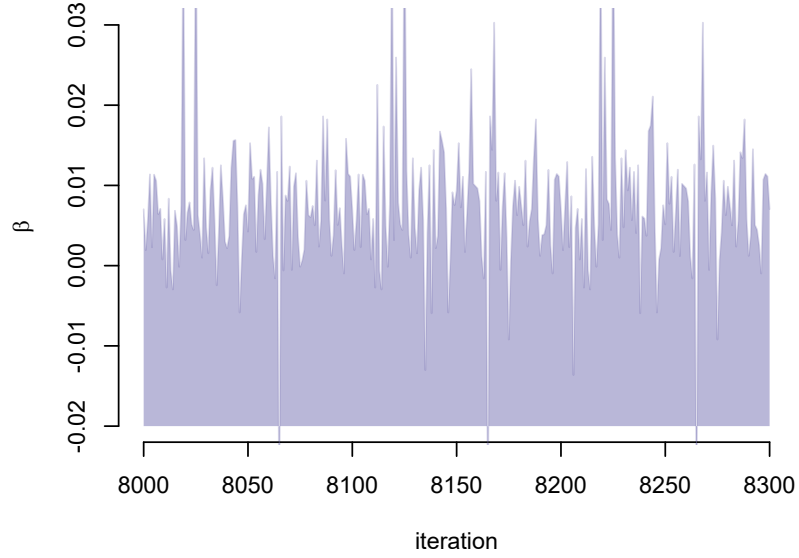
$$\theta_k|\theta_{k-1} \sim \mathcal{N}\left(\theta_k; \theta_{k-1} + \frac{1}{2}\mathcal{H}^{-1}(\theta_{k-1})\mathcal{G}(\theta_{k-1}), \mathcal{H}^{-1}(\theta_{k-1})\right).$$

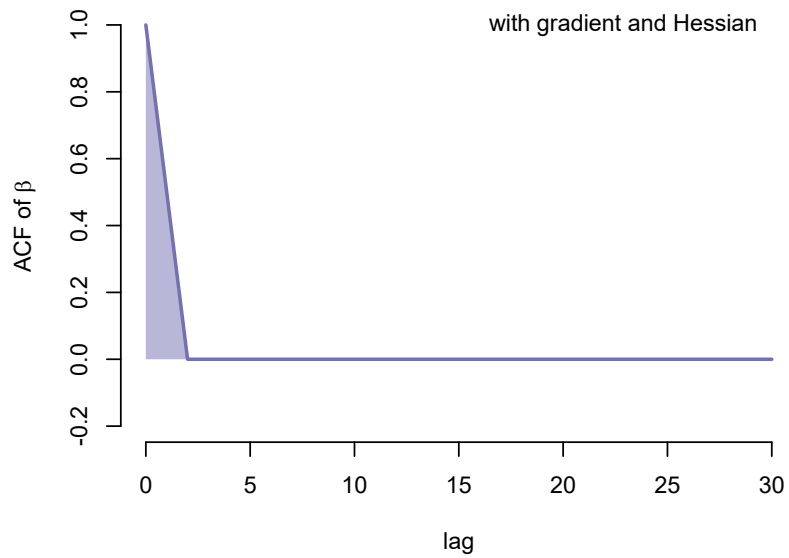
UPPSALA UNIVERSITET

Better informed proposals [III/III]: Philips curve









UPPSALA UNIVERSITET

What are some open questions?

- Decreasing computational time when T is large.

  Correlating and improving the particle filter.

- Obtaining better mixing when $p = |\theta|$ is large (>5).

  Add gradient and Hessian information into proposal.

- Devising better mixing when $n_x = |x|$ is large (>10).

  Improving the particle filter.

- Decrease the amount of tuning by the user.

  Adaptive algorithms and rules-of-thumb.

What did we do?

- Gave a (hopefully) gentle introduction to (P)MCMC.
- Developed some intuition for PMH and its pros/cons.
- Discussed some recent developments.

Why did we do this?

- PMH is general algorithm for Bayesian inference in SSMs.
- Relatively simple to implement and tune.

What are you going to do now?

- Remember that the PMH algorithm exist.
- Learning more by reading our tutorial.
- Try to implement the algorithm yourself.

UPPSALA
UNIVERSITET

Getting started with particle Metropolis-Hastings for inference in nonlinear dynamical models

Johan Dahlin* and Thomas B. Schön†

April 1, 2016

Abstract

We provide a gentle introduction to the particle Metropolis-Hastings (PMH) algorithm for parameter inference in nonlinear state space models (SSMs) together with a software implementation in the statistical programming language R. Throughout this tutorial, we develop an implementation of the PMH algorithm (and the integrated particle filter), which is distributed as the package **pmhtutorial** available from the CRAN repository. Moreover, we provide the reader with some intuition for how the algorithm operates and discuss some solutions to numerical problems that might occur in

Complete tutorial on PMH is available at arXiv:1511.01707.

Thank you for listening

Comments, suggestions and/or questions?

Johan Dahlin
uu@johandahlin.com
research.johandahlin.com

Remember: the tutorial is available at arXiv:1511.01707

UPPSALA UNIVERSITET

Particle filtering [I/II]

An instance of sequential Monte Carlo (SMC) samplers.

Estimates $\mathbb{E}\left[\varphi(x_t)|y_{1:t}\right]$ and $p_\theta(y_{1:T})$.

Computational cost of order $\mathcal{O}(NT)$ (with $N \sim T$).

Well-understood statistical properties.

(unbiasedness, large deviation inequalities, CLTs)

References:

A. Doucet and A. Johansen, A tutorial on particle filtering and smoothing. In D. Crisan and B. Rozovsky (editors), The Oxford Handbook of Nonlinear Filtering. Oxford University Press, 2011.

O. Cappé, S.J. Godsill and E. Moulines, An overview of existing methods and recent advances in sequential Monte Carlo. In Proceedings of the IEEE 95(5), 2007.

UPPSALA UNIVERSITET

Particle filtering [II/II]

By iterating:

Resampling: $\mathbb{P}(a_t^{(i)} = j) = \widetilde{w}_{t-1}^{(j)}$, for $i, j = 1, \ldots, N$.

Propagation: $x_t^{(i)} \sim f_\theta\left(x_t \middle| x_{t-1}^{a_t^{(i)}}\right)$, for $i = 1, \ldots, N$.

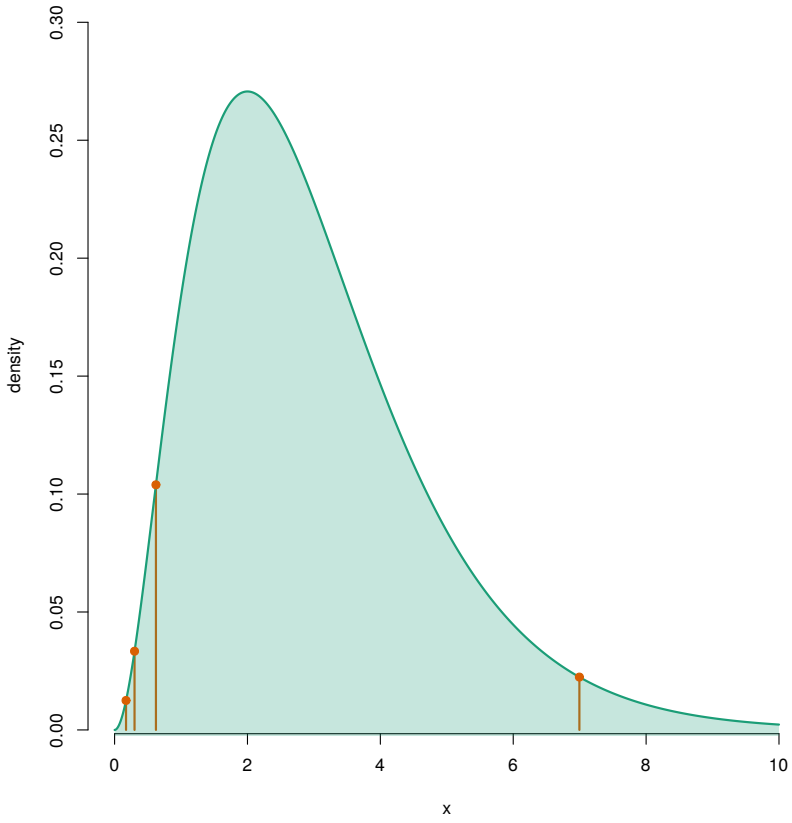
Weighting: $w_t^{(i)} = g_\theta\left(y_t \middle| x_t^{(i)}\right)$, for $i = 1, \ldots, N$.

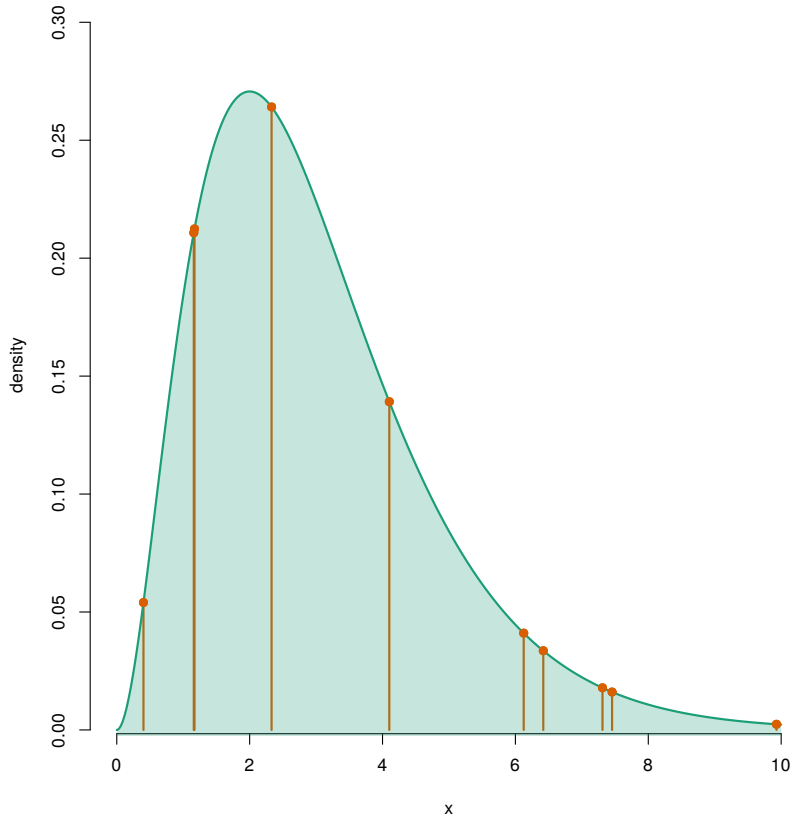
We obtain the particle system

$$\left[x_{0:T}^{(i)}, w_{0:T}^{(i)}\right]_{i=1}^{N}.$$

UPPSALA UNIVERSITET

Particle filtering: state estimation

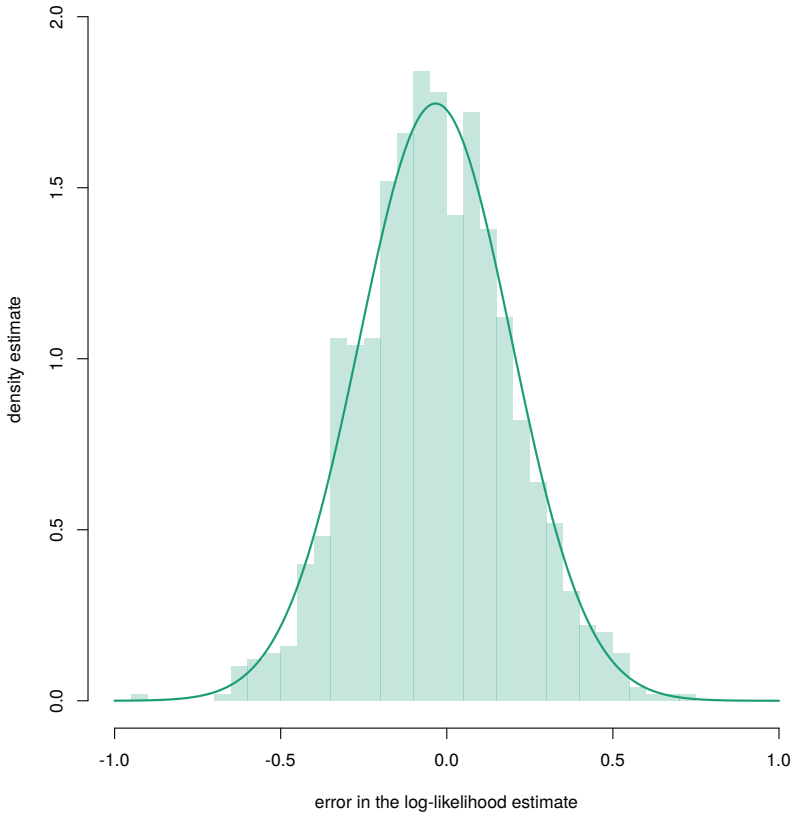


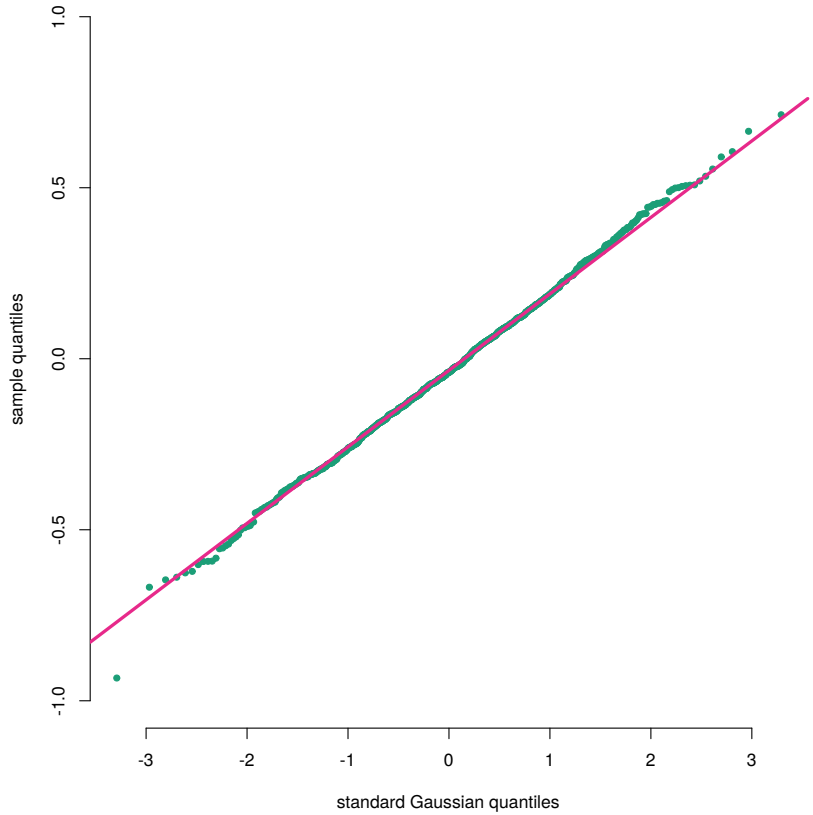


$$\widehat{\varphi}_t^N \triangleq \widehat{\mathbb{E}}\Big[\varphi(x_t)|y_{1:t}\Big] = \sum_{i=1}^{N} w_t^{(i)} \varphi\big(x_t^{(i)}\big), \qquad \sqrt{N}\Big(\varphi_t - \widehat{\varphi}_t^N\Big) \xrightarrow{d} \mathcal{N}\Big(0, \sigma_t^2(\varphi)\Big).$$

UPPSALA UNIVERSITET

Particle filtering: likelihood estimation

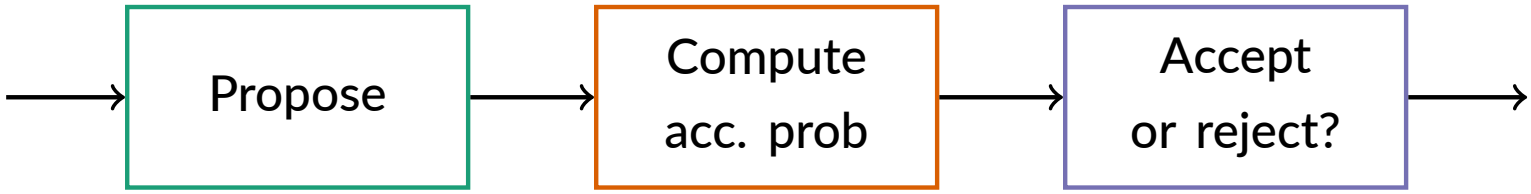




$$\underbrace{\log \widehat{p}_\theta(y_{1:T})}_{\triangleq \widehat{\ell}(\theta)} = \sum_{t=1}^{T} \log\left(\sum_{i=1}^{N} w_t^{(i)}\right) - T\log N, \quad \sqrt{N}\left(\ell(\theta) - \widehat{\ell}(\theta) + \frac{\sigma^2_{\widehat{\pi}}}{2N}\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\widehat{\pi}}\right).$$

UPPSALA UNIVERSITET

Particle Metropolis-Hastings [I/III]



- Propose: $\theta' \sim q(\theta'|\theta_{k-1}, u')$ and $u' \sim \mathrm{PF}(\theta')$.
- Compute $\widehat{p}_{\theta'}(y_{1:T}|u')$ and the acceptance probability:

$$\alpha(\theta', \theta_{k-1}) = 1 \wedge \frac{p(\theta')}{p(\theta_{k-1})} \frac{\widehat{p}_{\theta'}(y_{1:T}|u')}{\widehat{p}_{\theta_{k-1}}(y_{1:T}|u_{k-1})} \frac{q(\theta_{k-1}|\theta', u')}{q(\theta'|\theta_{k-1}, u_{k-1})}.$$

- Accept or reject? $\theta' \to \theta_k$ and $u' \to u_k$ w.p. $\alpha(\theta', \theta_{k-1})$.

UPPSALA UNIVERSITET

Particle Metropolis-Hastings [II/III]

The target distribution is given by the parameter proposal

$$\pi(\theta) = \frac{p_\theta(y_{1:T})p(\theta)}{p(y_{1:T})}.$$

An unbiased estimator of the likelihood is given by

$$\mathbb{E}_m\left[\widehat{p}_\theta(y_{1:T}|u)\right] = \int \widehat{p}_\theta(y_{1:T}|u)m_\theta(u)\,\mathrm{d}u = p_\theta(y_{1:T}).$$

An extended target is given by

$$\pi(\theta, u) = \frac{\widehat{p}_\theta(y_{1:T}|u)m_\theta(u)p(\theta)}{p(y_{1:T})} = \frac{\widehat{p}_\theta(y_{1:T}|u)m_\theta(u)\pi(\theta)}{p_\theta(y_{1:T})}.$$

UPPSALA UNIVERSITET

Particle Metropolis-Hastings [III/III]

$$
\begin{aligned}
\int \pi(\theta, u) \, \mathrm{d}u &= \int \frac{\widehat{p}_\theta(y_{1:T}|u) m_\theta(u) \pi(\theta)}{p_\theta(y_{1:T})} \, \mathrm{d}u \\
&= \frac{\pi(\theta)}{p_\theta(y_{1:T})} \underbrace{\int \widehat{p}_\theta(y_{1:T}|u) m_\theta(u) \, \mathrm{d}u}_{=p_\theta(y_{1:T})}, \\
&= \pi(\theta).
\end{aligned}
$$

That is, the marginal is the desired target distribution and the Markov chain is kept invariant.