

ENM 531: Data-driven Modeling and Probabilistic Scientific Computing

Lecture #14: Markov Chain Monte Carlo

Paris Perdikaris
March 9, 2021



The Metropolis algorithm

- ▶ Choose a symmetric proposal matrix Q . So, $Q_{ab} = Q_{ba}$.
- ▶ Initialize $x_0 \in X$.
- ▶ for $i \in 0, 1, 2, \dots, n - 1$:
 - ▶ Sample proposal x from $Q(x_i, x)$ if x is discrete, otherwise, $p(x | x_i)$.
 - ▶ Sample r from $\text{Uniform}(0, 1)$.
 - ▶ If

$$r < \frac{\tilde{\pi}(x)}{\tilde{\pi}(x_i)},$$

accept and $x_{i+1} = x$.

- ▶ Otherwise, reject and $x_{i+1} = x_i$.

Output: x_0, x_1, \dots, x_n

- ▶ Symmetric proposals include:

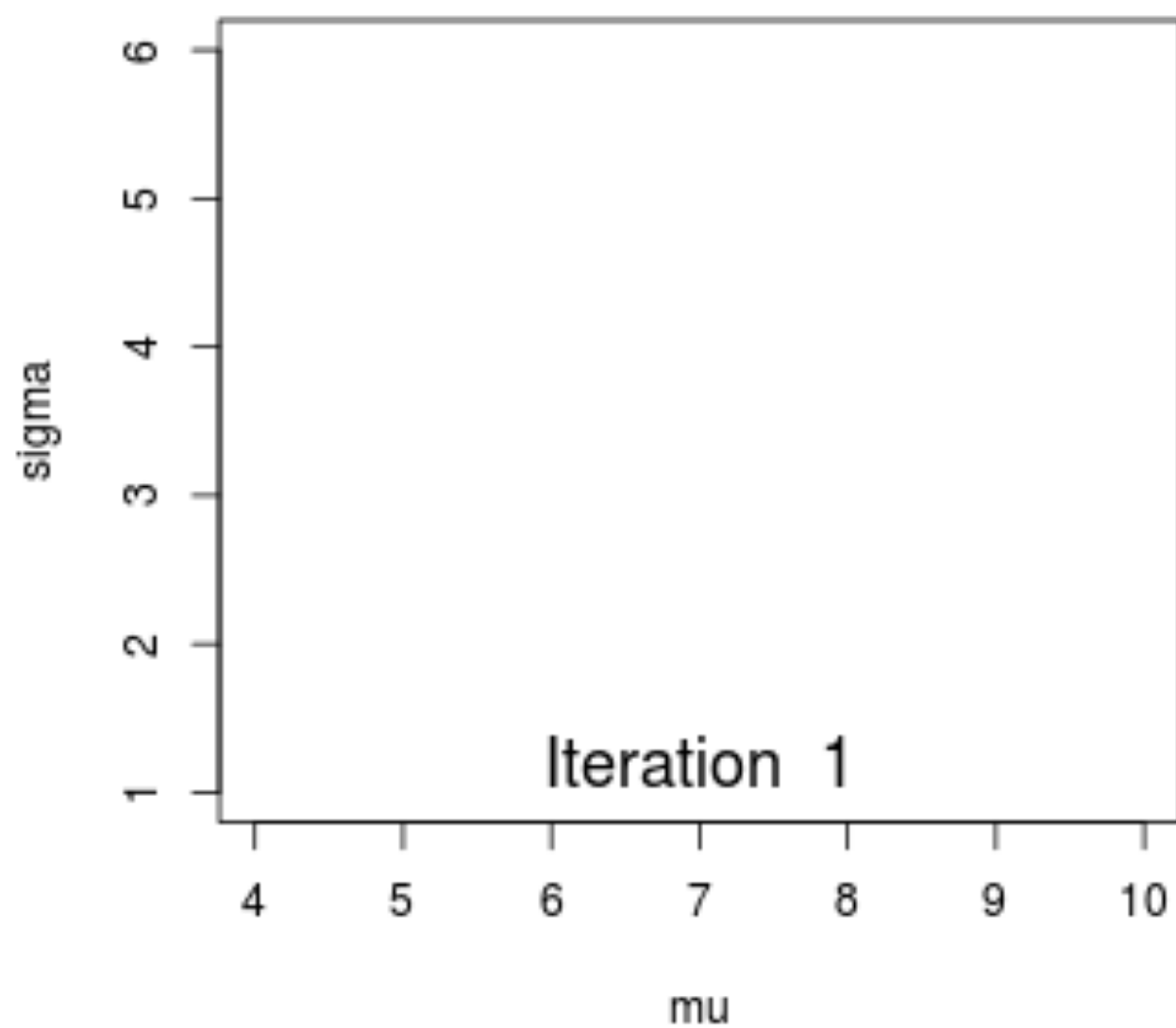
$$J(\theta^* | \theta^{(s)}) = \text{Uniform}(\theta^{(s)} - \delta, \theta^{(s)} + \delta)$$

and

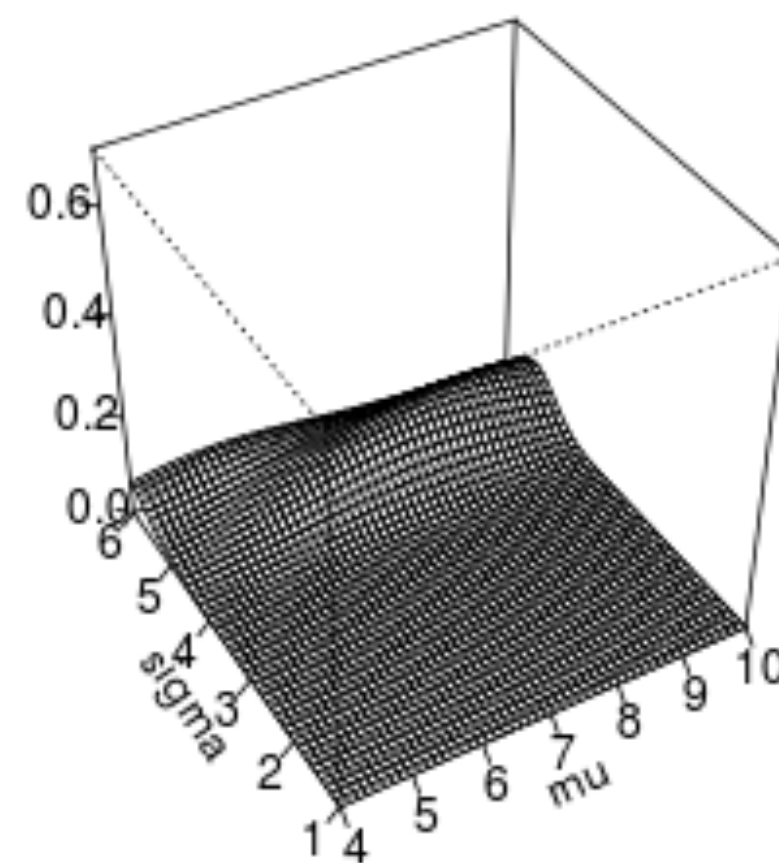
$$J(\theta^* | \theta^{(s)}) = \text{Normal}(\theta^{(s)}, \delta^2).$$

The Metropolis algorithm

Markov chains



Posterior density



The Metropolis algorithm for Bayesian inference

Goal: We want to sample from

$$p(\theta \mid y) = \frac{f(y \mid \theta)\pi(\theta)}{m(y)}.$$

Typically, we don't know $m(y)$.

The notation is a bit more complicated, but the set up is the same.

We'll approach it a bit differently, but the idea is exactly the same.

We know $\pi(\theta)$ and $f(y \mid \theta)$, so we can draw samples from these.

Our notation here will be that we assume parameter values $\theta_1, \theta_2, \dots, \theta_s$ which are drawn from $\pi(\theta)$.

We assume a new parameter value comes in that is θ^* .

The Metropolis algorithm for Bayesian inference

The Metropolis algorithm proceeds as follows:

1. Sample $\theta^* \sim J(\theta \mid \theta^{(s)})$.
2. Compute the acceptance ratio (r):

$$r = \frac{p(\theta^* | y)}{p(\theta^{(s)} | y)} = \frac{p(y \mid \theta^*)p(\theta^*)}{p(y \mid \theta^{(s)})p(\theta^{(s)})}.$$

3. Let

$$\theta^{(s+1)} = \begin{cases} \theta^* & \text{with prob } \min(r, 1) \\ \theta^{(s)} & \text{otherwise.} \end{cases}$$

Remark: Step 3 can be accomplished by sampling $u \sim \text{Uniform}(0, 1)$ and setting $\theta^{(s+1)} = \theta^*$ if $u < r$ and setting $\theta^{(s+1)} = \theta^{(s)}$ otherwise.

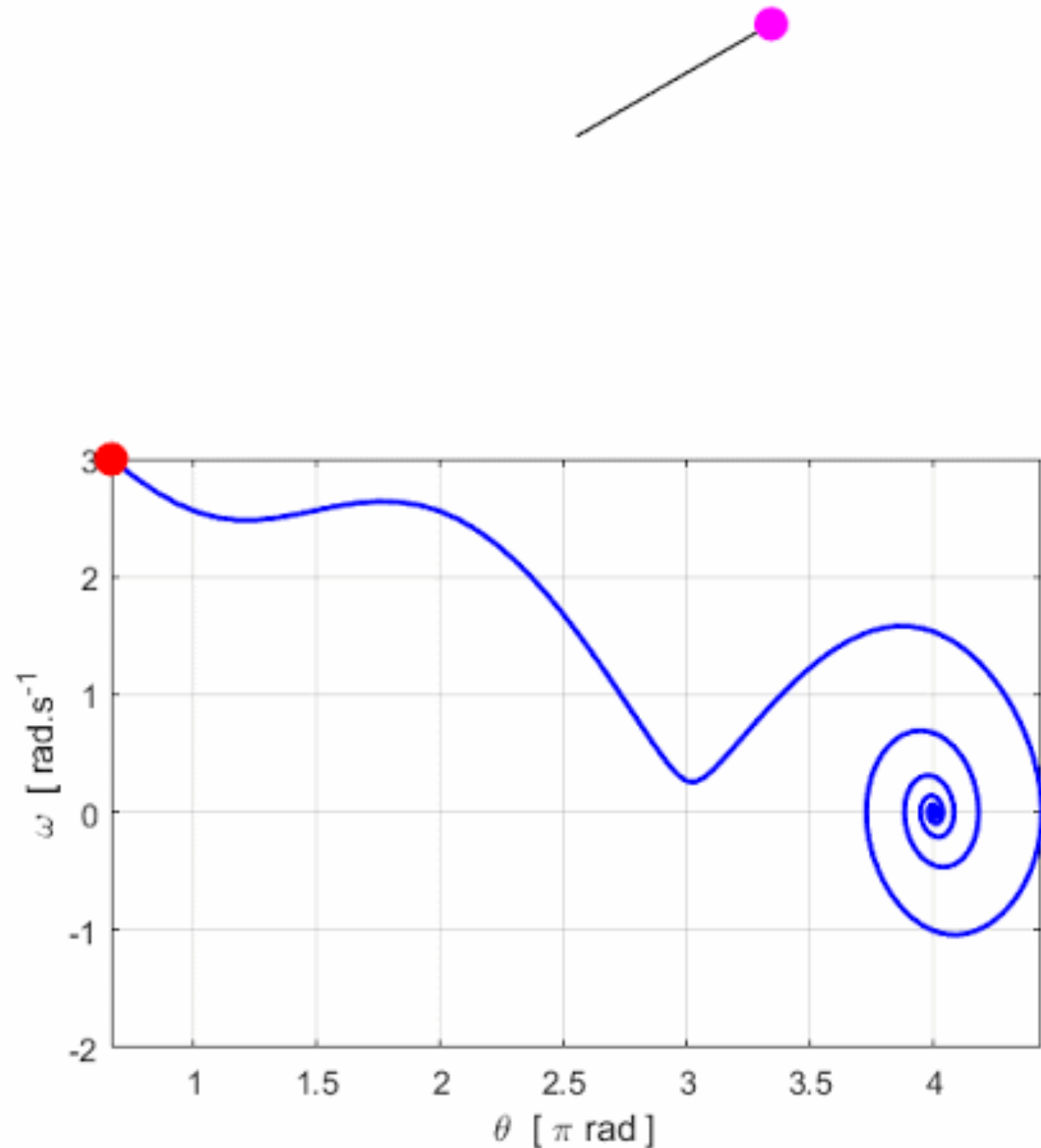
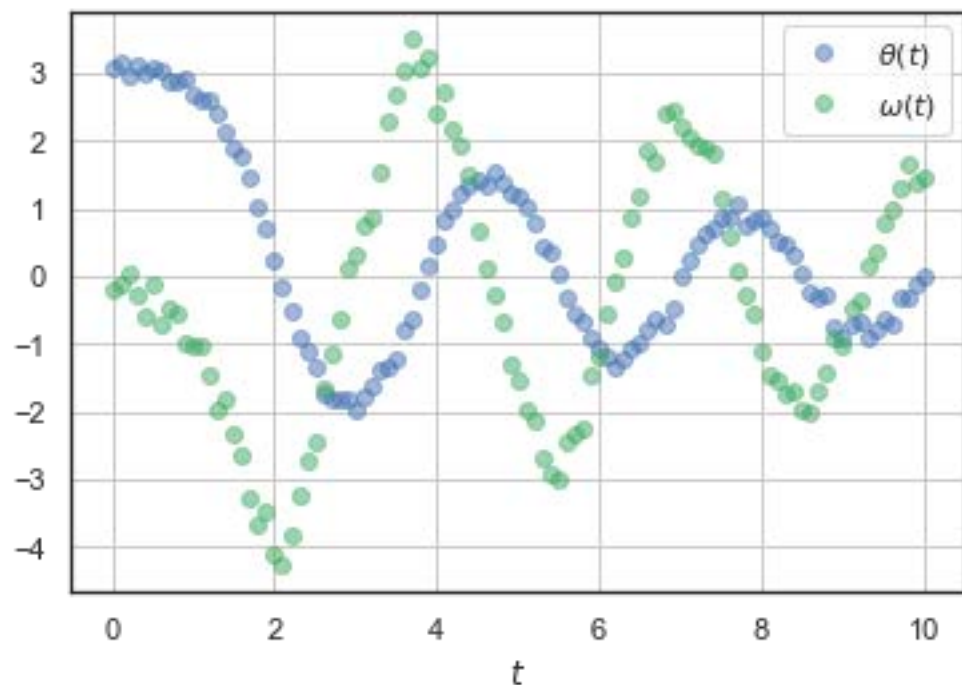
Bayesian calibration of dynamical systems

Example: Damped pendulum

$$\begin{aligned}\frac{d\theta}{dt} &= \omega, \\ \frac{d\omega}{dt} &= -b\omega - c \sin(\theta).\end{aligned}$$

Given some noisy time-series data

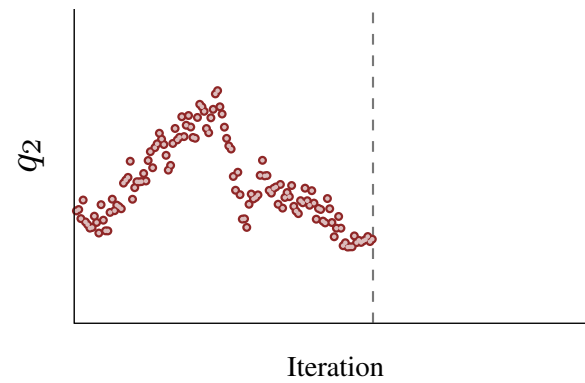
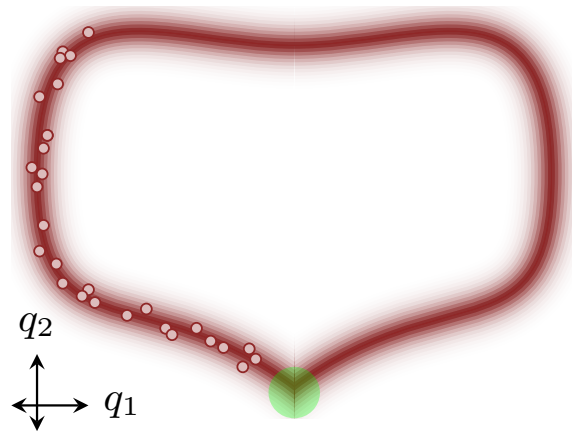
$$\mathcal{D} := \{\theta(t_i), \omega_i(t_i)\}, \quad i = 1, \dots, n$$



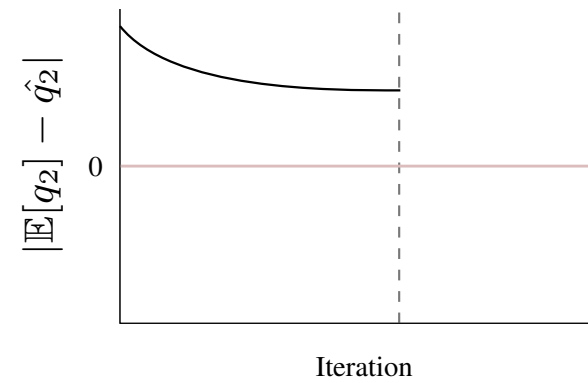
Infer a posterior distribution over the unknown parameters

$$p(b, c | \mathcal{D})$$

Hamiltonian Monte Carlo

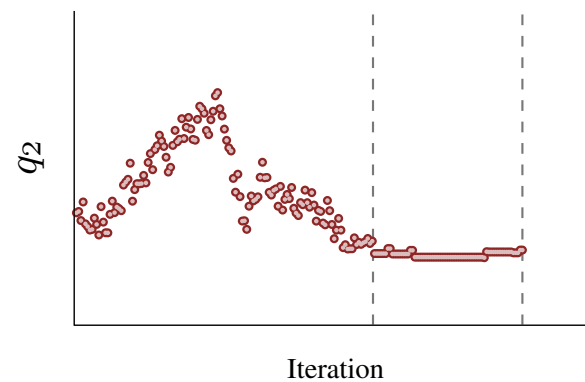
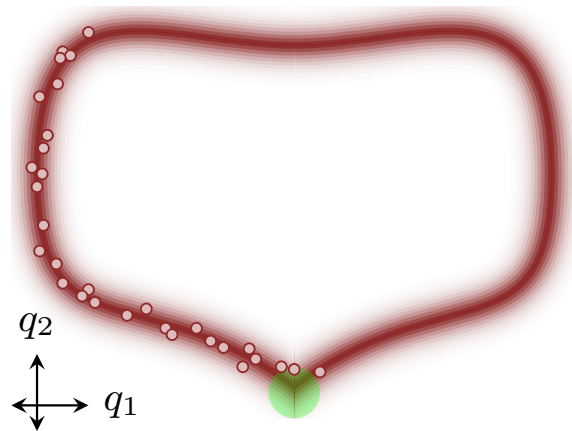


(a)

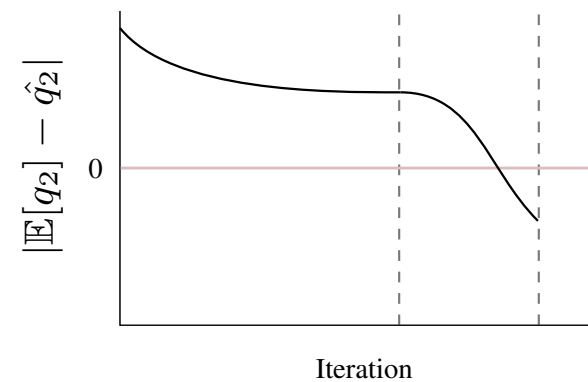


In practice, pathological regions of the typical set usually cause Markov chains to get “stuck”.

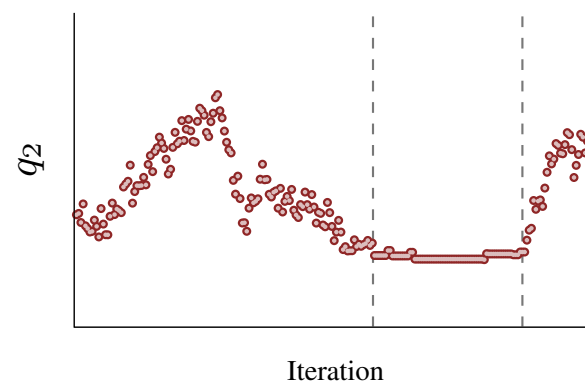
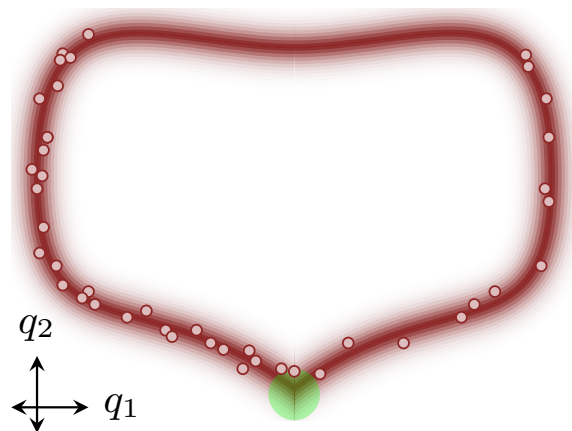
(a) Initially the Markov chain explores well-behaved regions of the typical set, avoiding the pathological neighborhood entirely and biasing Markov chain Monte Carlo estimators.



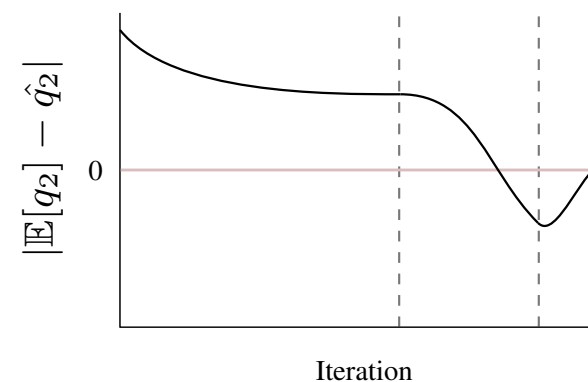
(b)



(b) If the Markov chain is run long enough then it will get stuck near the boundary of the pathological region, slowly correcting the Markov chain Monte Carlo estimators.

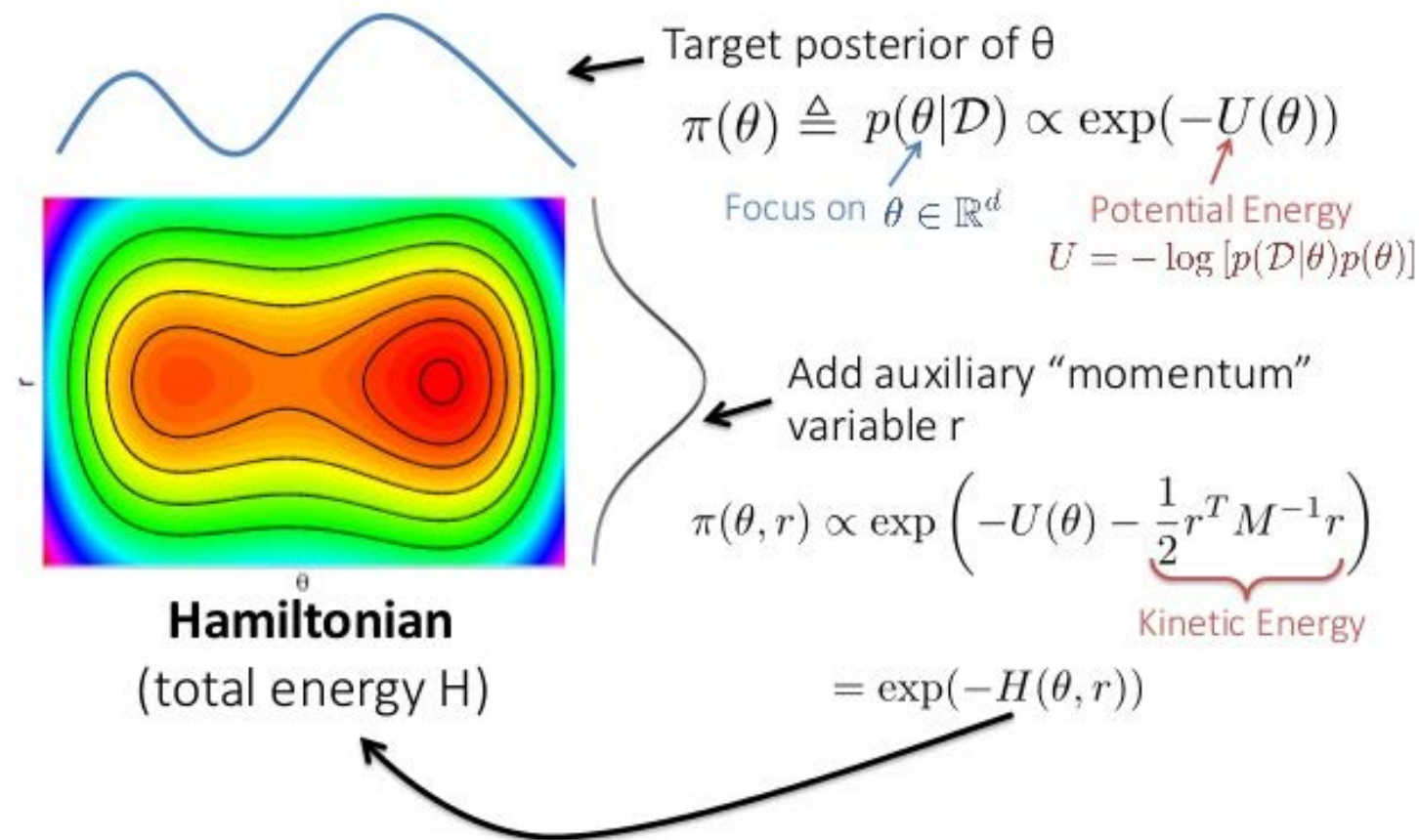
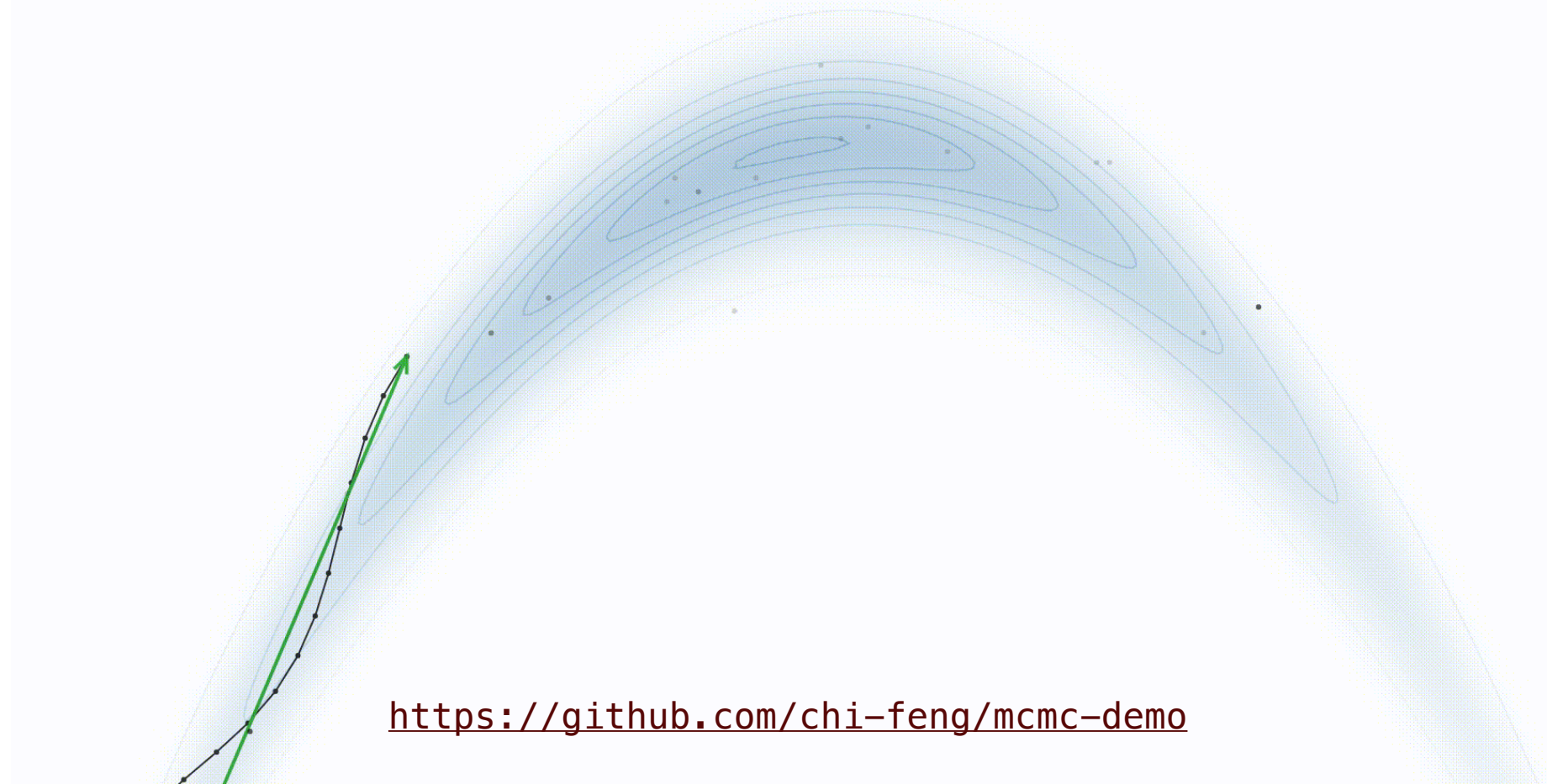


(c)

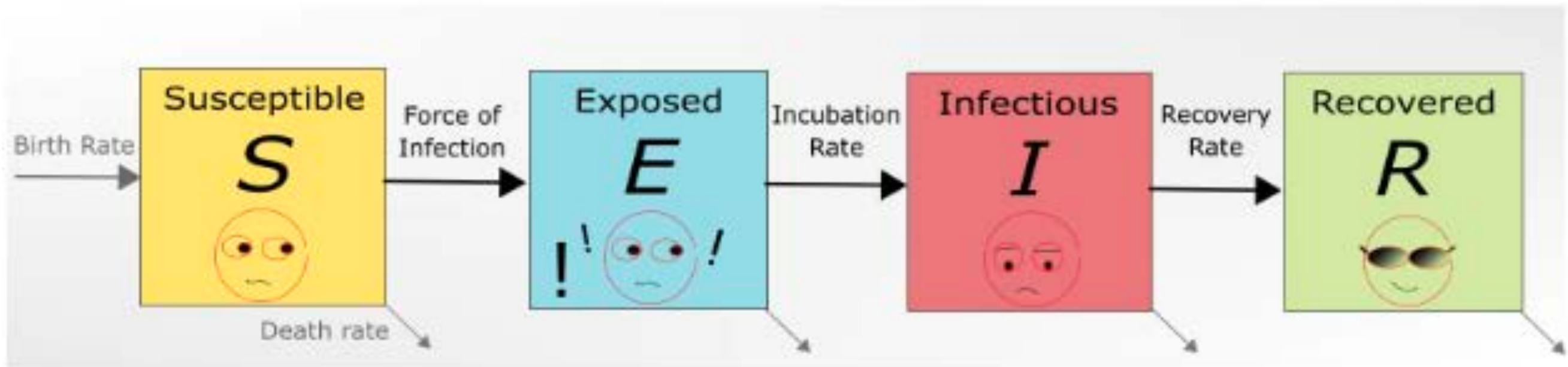


(c) Eventually the Markov chain escapes to explore the rest of the typical set. This process repeats, causing the resulting estimators to oscillate around the true expectations and inducing strong biases unless the chain is improbably stopped at exactly the right time.

Hamiltonian Monte Carlo



Example: Epidemiology models



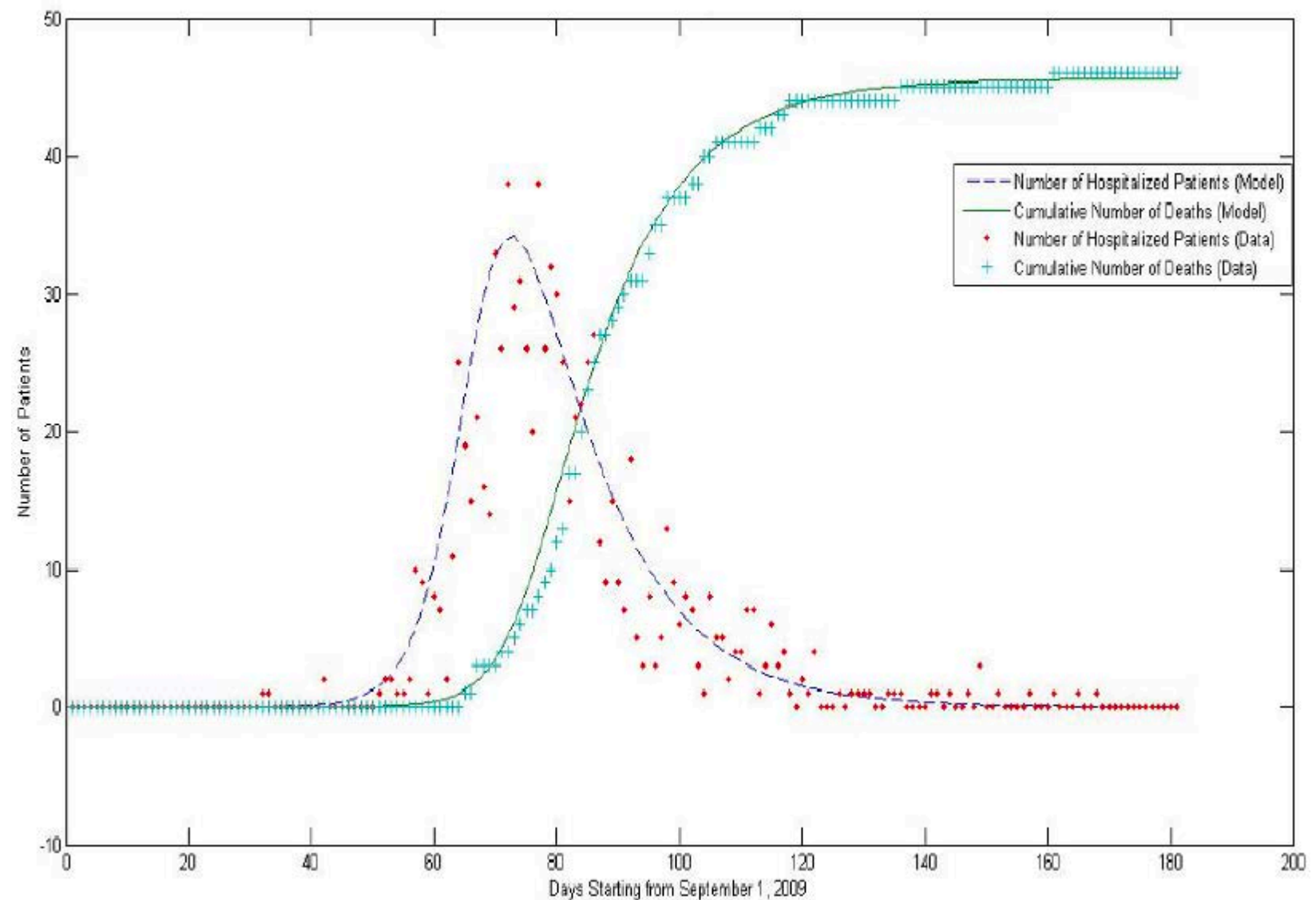
$$\frac{dS}{dt} = -\frac{\beta SI}{N}$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$

$$\frac{dI}{dt} = \sigma E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

$$N = S + E + I + R$$



Probabilistic programming

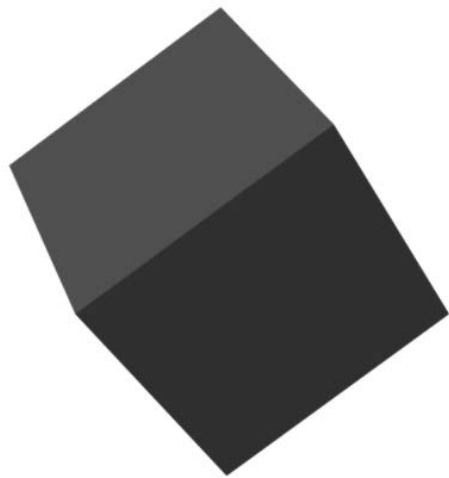


<http://mc-stan.org/>



<https://github.com/pymc-devs/pymc3>

Edward



<http://edwardlib.org/>



<https://github.com/uber/pyro>