Computational Statistics & Machine Learning

Lecture 15

Parallel Tempering (Replica Exchange)

Mark Girolami
mag92@cam.ac.uk

Department of Engineering
University of Cambridge

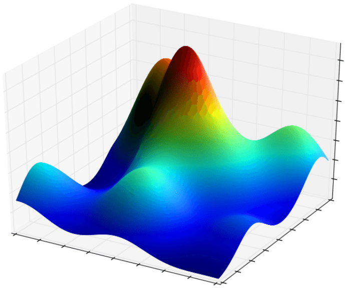
October 3, 2025

Overview

- Motivation: Multimodal targets
- Parallel tempering MCMC method
- Examples of the MCMC schemes

Multiple modes

MCMC methods perform well when the target π is uni-modal but struggle to explore the space for multi-modal targets.



This is an important problem: It arises in

- Machine learning
- Protein folding
- Nonconvex optimization of high-dimensional landscapes

How can we explore multiple modes?

The idea:

- Define tempered (smoothed) versions of the π
- Run multiple MCMC chains for each tempered distribution and combine.

This will allow us to explore efficiently and will make transition between modes easier.

How can we explore multiple modes?

Let $U(x) = -\log \pi(x)$ (also called *the potential*). Then define

$$\pi_\beta(x) = \pi(x)^\beta = \exp(-\beta U(x)). \tag{1}$$

- β is called inverse temperature, i.e., $\beta = 1/T$ (equivalent formulation).
- For each β, we obtain a "smoothed" version of π.
- Exploring smoothed densities might be way easier in terms of getting to "difficult to travel" regions otherwise.

How can we explore multiple modes?

Parallel tempering: Sample from "each" replica and swap them with a certain probability.

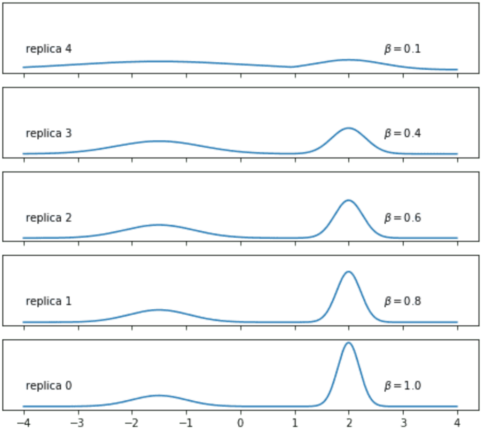


Figure: photo: https://www.tweag.io/blog/2020-10-28-mcmc-intro-4/

Replica exchange

We can use any valid MCMC scheme to sample from each replica (π_β).

- Each chain will explore a tempered version of the target using an MCMC scheme if not involved in a swap.
- We can then use these "replicas" as proposals for each other
  - Replica exchange.
- Only two chains are involved in a swap at a single iteration (other chains run independently during a swap)

Replica exchange

Let $(X_k^i)_{k \geq 0}$ be the Markov chain generated at ith chain with stationary measure π_{β_i} for K different replicas.

At iteration $t + 1$

- Choose $i, j \in \{1, \ldots, K\}$ for a swap.
- Run other chains with any valid MCMC scheme.
- Compute the acceptance probability for the product measure $\pi_{\beta_i} \times \pi_{\beta_j}$

  $$\alpha = \min \left\{ 1, \frac{\pi_{\beta_i}(x_t^j) \times \pi_{\beta_j}(x_t^i)}{\pi_{\beta_i}(x_t^i) \times \pi_{\beta_j}(x_t^j)} \right\}$$

  This is the MH probability for a proposed sample (x_t^j, x_t^i) for a chain at the state (x_t^i, x_t^j).
- This will leave each chain's marginal invariant.

Replica exchange

Swap strategy has to be well chosen.

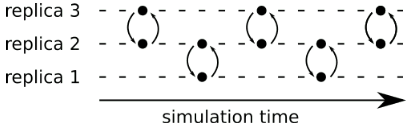


Figure: A swap strategy (Fig. from the same webpage – needs to be replaced).