Parallel Tempering MCMC Examples

Baris Alparslan

October 2025

1 Results and Examples

Coupled Markov chain approach is usually useful when there is multi-modality in the data where chain is likely to stuck at some state. In this report, I will elaborate the discussion on parallel tampering with examples. Examples will focus on mimicking the multi-modal data sets, and exploring parameter space with MCMC methods. For the sake of simplicity, I considered symmetric proposal distributions. Additionally, prior distributions are selected such that they have minimal effect on the selection process. Also, please note that all the simulations use 1×10^5 iterations in total. For the plotting, I considered first %80 of the samples as burn-in period.

1.1 Example 0

This example is direct implementation of the example from [Wilkinson, 2013]. To remember the setting, we have double potential well distribution ($\pi(x)$) with potential distribution as

$$U(x) = \gamma(x^2 - 1)^2$$

and $\pi(x) = \exp(-U(x))$.

The overall aim is to estimate γ given data. However, the issue is that, regular MCMC model may fail due to bi-modality (See figure 1).

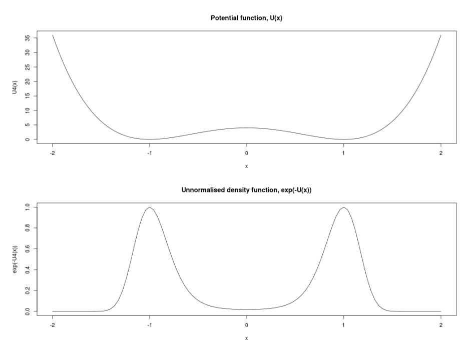

Figure 1: Potential function ($U(x)$) and Density plot ($\pi(x)$) when $\gamma = 4$

In fact, we observe that the higher true value of γ gets, the more difficult it gets for chain to explore whole space. Additionally, chain mixes very poorly for different values of (See figure 2).

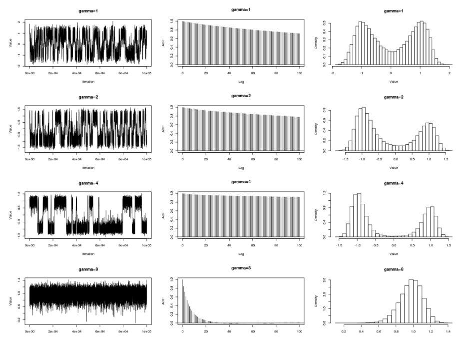

Figure 2: Sampling performance of regular MH sampling

We can use coupled chain approach where we swap states based on the temperature of the chains. Note that temperatures here are $1/\gamma$ since gamma takes values greater than 1. One can find acceptance probability for state swaps easily using the same formulation. The swapping structure allows chain to mix and explore better (See figure 3).

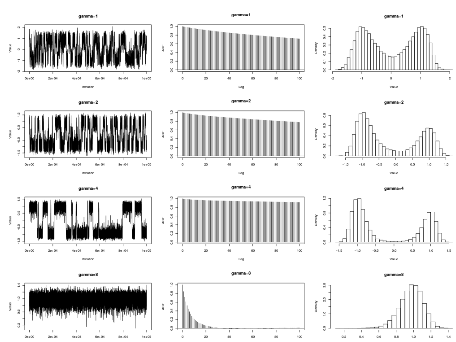

Figure 3: Sampling performance coupled chain with state swaps (Parallel tampering)

1.2 Example 1

Now, we consider a slightly different example with bi-modality. Suppose we have,

$$X = \begin{cases} Z_1 & \text{w.p.} \quad 0.5 \\ Z_2 & \text{w.p.} \quad 0.5 \end{cases}$$

where $Z_1 \sim N(\theta_1, \sigma^2) \quad Z_2 \sim N(\theta_2, \sigma^2)$. The task is estimating θ_1 and θ_2 given σ^2 is known. We consider true values as $(\theta_1, \theta_2) = (6, -6)$ and $\sigma^2 = 1$. Then, the bimodality is obvious in figure 4.

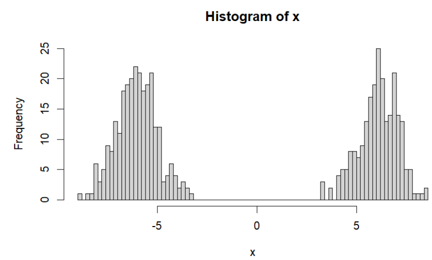

Figure 4: Density for given Gaussian mixture example

For simulation purpose, we consider symmetric random walk proposals for parameters as follows. σ^2_{tune} is tuned to obtain clear visuals. Additionally, we take normal non-informative (high variance) priors.

$$\theta_1 \sim N(0, 100), \quad \theta_2 \sim N(0, 100)$$
$$\theta_1^{(i+1)} \sim N(\theta_1^{(i)}, \sigma^2_{\text{tune}})$$
$$\theta_2^{(i+1)} \sim N(\theta_2^{(i)}, \sigma^2_{\text{tune}})$$

Given these model specifications, the acceptance ratio can be described as

$$A = \min\left(1, \frac{p\left(X|\theta_1^{(i+1)}, \theta_2^{(i+1)}\right) p_{\theta_1}(\theta_1^{(i+1)}) p_{\theta_2}(\theta_2^{(i+1)})}{p\left(X|\theta_1^{(i)}, \theta_2^{(i)}\right) p_{\theta_1}(\theta_1^{(i)}) p_{\theta_2}(\theta_2^{(i)})}\right)$$

As we have discussed before, we expect that single chain MH algorithm would fail to mix and estimate in this example. In fact, when we check the figure 5, we observe that chain is sticky and the convergence is not obtained as chain seems to stuck at opposite location.

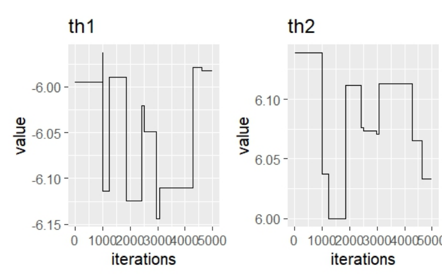

Figure 5: Sampling performance of MH algorithm

After running the coupled chain with coupled model with heat parameters as $(0.1, 0.3, 0.5, 0.7, 1)$ for 5 different chains, we observe the following improvement in the performance (See figure 6). Note that this is only for state swaps, not heat transfer. Also, only chain 5 is cold chain, where we should only consider cold chain for estimation purposes.

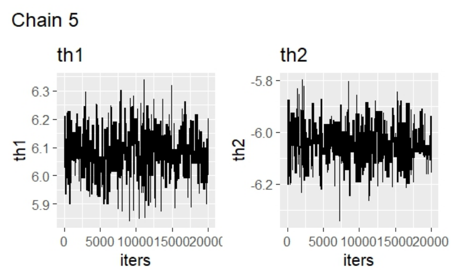

Figure 6: Sampling performance of coupled chain (Parallel tampering) (cold chain with state swaps)

1.3 Example 2

This example is useful to see the difference between coupled chain and regular MH algorithm because the multi-modal structure makes it even harder for regular MH algorithm to explore. The setting is directly coming from [Bob Verity, 2024]. In detail, we have

$$X \overset{i.i.d}{\sim} N(\alpha^2\beta + \epsilon, 1), \quad \epsilon \sim N(0, 1)$$

where negative and positive alpha values give same density which leads multi-modality. Only, α and β (with true values being $\alpha = 3, \beta = 2$) are sampled with following priors.

$$\alpha \sim \text{Unif(-10,10)} \quad \beta \sim \text{Unif(0,10)}$$

The regular MH chain with a similar acceptance ratio as before results in following samples (after burn-in, see figure 7). It is obvious that chain doesn't mix well, also chain doesn't converge to the true value.

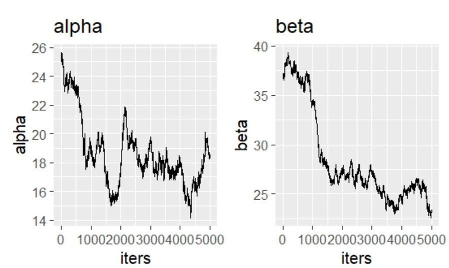

Figure 7: Sampling performance of regular MH chain

When we conduct the same experiment with coupled chain with state swaps where we have 5 chain with heat parameters as $(0.1, 0.3, 0.5, 0.7, 1)$, we observe a clear improvement in the mixing and convergence behavior of the chain (See figure 8). Note that these are samples from cold chain.

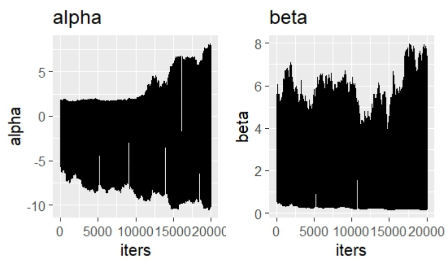

Figure 8: Sampling performance of parallel tampering (cold chain with state swaps)

References

[Bob Verity, 2024] Bob Verity, P. W. (2024). Parallel tempering. https://mrc-ide.github.io/drjacoby/articles/metropolis_coupling.html. Accessed: 10/12/2025.

[Wilkinson, 2013] Wilkinson, D. (2013). Parallel tempering and metropolis coupled mcmc. darrenjw.wordpress.com/2013/09/29/parallel-tempering-and-metropolis-coupled-mcmc/. Accessed: 10/12/2025.