Some insights into the reliability of Adaptive Markov chain Monte Carlo

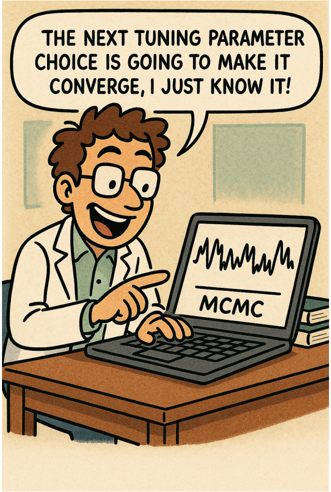
Austin Brown

Texas A&M University

September 27, 2025

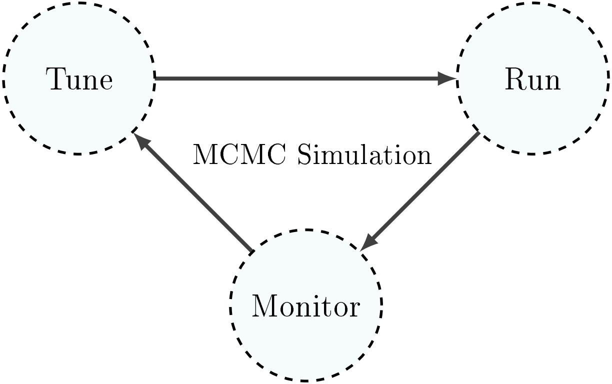
Ideal setup

Choose a "good" tuning parameter γ depending on information from the target, and simulate samples from a Markov process until the marginal samples *stabilize* to samples from π.



Tuning MCMC is hard

- **Problem:** Tuning an MCMC algorithm is incredibly difficult and trial and error is computationally expensive and wasteful.



How I think of Adaptive MCMC

Try to make new MCMC algorithms that are easy to implement for scientists not in this room without access to optimal tuning parameters.

Adaptive MCMC

Choose an adaptation plan $\mathcal{Q} \equiv (\mathcal{Q}_t)_t$ (family of kernels) for updating the tuning parameter using the history.

1. Sample γ_{t+1} | history.
2. Sample state space $X_{t+1}|\gamma_{t+1}, X_t$

Examples. RWM adapting the covariance with the previous history. A covariance using information with the target would be ideal, but may not be readily available.

Convergence in Adaptive Markov chain Monte Carlo

We need to burn-in the tuning parameters up to T and then continue running t more times to converge as we would normally

$$T + t$$

Asymptotic results here [Roberts and Rosenthal, 2007].

Motivation

How to design “good" adaptive algorithms ("good" adaptation plans)? Are there optimal adaptation plans? etc.

Motivation

Metropolis-Hastings with exponential target and independent proposal

$$\gamma \exp(-\gamma x).$$

- If we adapt with $\gamma < 1$, can expect geometric convergence
- If $\gamma > 1$ (can't be too large), can expect polynomial convergence

So we likely have a phase transition in the convergence of adaptive MCMC when $\gamma = 1$ because of the **tail behavior change**.

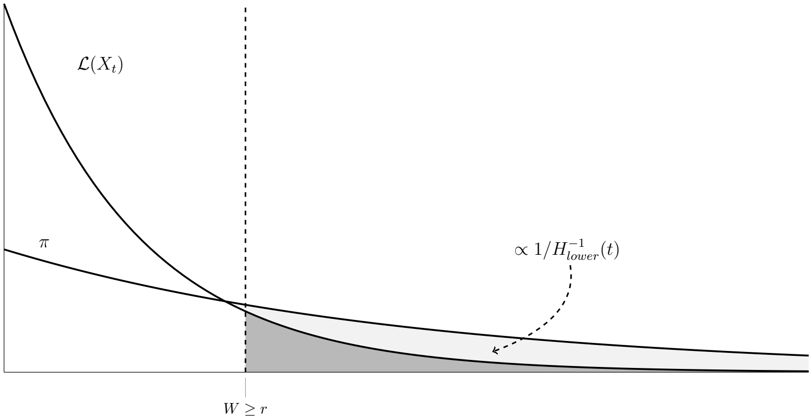
Takeaways from this talk

Adapting a Markov process may not improve the tail behavior enough, and the convergence can behave like a non-adapted Markov process with a potentially poor tuning parameter choice.

- Previous results eluded to this in specific examples [Schmidler and Woodard, 2011].
- Many adaptive algorithms are designed by adapting only on a compact set.

Intuition: tail mismatch

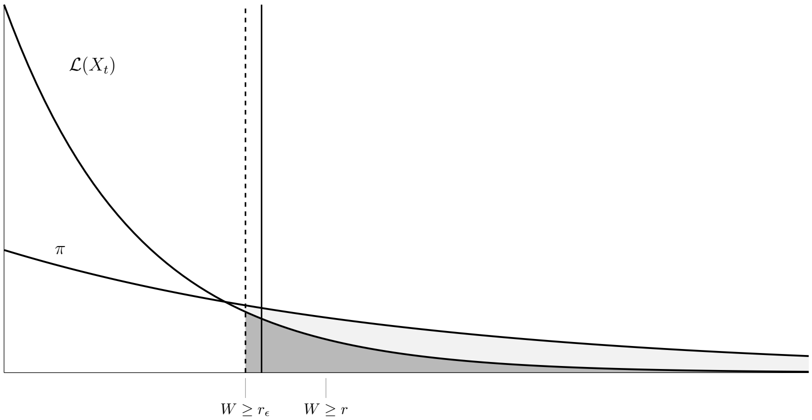
Find a function $W \geq 0$ that is not integrable with π but integrable with the adaptive process, then we have a tail mismatch:



$H_{lower}^{-1}(t)$ is often only good for slower than geometric rates like polynomial.

Intuition: tail mismatch can be robust to perturbation

The previous plot can be robust to small perturbations.



Intuition: robust tail mismatch implication

We can find a set A where the volume

$$\pi(A^\epsilon) \text{ differs from } \mathcal{L}(X_t)(A)$$

where A^ϵ is the $\epsilon-$inflation set.

Lower bound

Theorem (Outline)

Under this setting, for any adaptation plan $\mathcal{Q}$ that maintains this tail mismatch problem, we can find a δ

$$\frac{1}{H_{lower}^{-1}(t)} \lesssim \mathbb{P}_{\mathcal{Q}}(\|X_t - Y\| > \delta)$$

where H_{lower}^{-1}.

Upper bounds

Conditions for upper bounds

We need some subgeometric drift and local coupling conditions and a quantiative diminishing adaptation condition [Roberts and Rosenthal, 2007]. An upper bound rate on the closeness of the distributions of

$$X_{t+1} \mid \gamma_{t+1}, X_t \text{ and } X_{t+1} \mid \gamma_t, X_t$$

uniformly in X_t.

Total variation control

Theorem (Outline)

Under this setting, if the adaptation is fast, for $\epsilon > 0$

$$\left\| \mathcal{L}(X_{2\log(t/\epsilon)+t}) - \pi \right\|_{TV} \lesssim \frac{\log(t/\epsilon)}{H_{upper}^{-1}(t)} + \epsilon$$

- Could replace this with a "Wasserstein distance", but the lower bound won't change.

Example: convergence characterization of adaptive RWM

More generally, the upper bound requires a balance of the "adaptation burn-in time" and the convergence of the underlying Markov process:

$$\left\|\mathcal{L}(X_{T(t,\epsilon)+t}) - \pi\right\|_{\mathrm{TV}} \lesssim \frac{T(t,\epsilon)}{H_{upper}^{-1}(t)} + \epsilon$$

- Should take into account information from the convergence of the Markov process with the adaptation strategy.

Example: convergence characterization of IMH

Proposition (Roughly)

If adaptation settles fast, adaptive IMH for the exponential target is polynomial

$$n^{-b}$$

with b in some range of values depending on the best in the lower bound and worst tuning parameter choices in the upper bound.

Note: Might be useful for characterizing when the CLT will not hold.

Example: convergence characterization of adaptive RWM

Proposition (Roughly)

If adaptation settles fast and for certain targets with heavier tails, adaptive RWM is subgeometric

$$\exp\left[-bt^a\right]$$

with b in some range of values depending on the best and worst tuning parameter choices.

References I

Gareth O. Roberts and Jeffrey S. Rosenthal. Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *Journal of Applied Probability*, 44(2):458–475, 2007.

Scott C. Schmidler and Dawn B. Woodard. Lower bounds on the convergence rates of adaptive MCMC methods. *Tech. rep., Duke Univ.*, 2011.