Advanced Methods in Markov chain Monte Carlo

Erik Thiede

Before we start…..

- Focus is on “big picture” of the algorithms
- Slides on the FWAM2 website have an annotated bibliography at the end.

Recap on Markov Chain Monte Carlo

Known exactly

$$\mu_f = \frac{\int f(x)\boxed{\rho(x)}dx}{\boxed{\int \rho(x)dx}}$$

Intractable.

Bayesian MCMC:

$$\rho(\theta) = p\left(y|\theta\right)p\left(\theta\right)$$

$$\int \pi(\theta)dx = \int_\theta p\left(y|\theta\right)p\left(\theta\right)d\theta = p(y)$$

Statistical Mechanics:

$$\rho(x) = e^{-H(x)/k_B T}$$

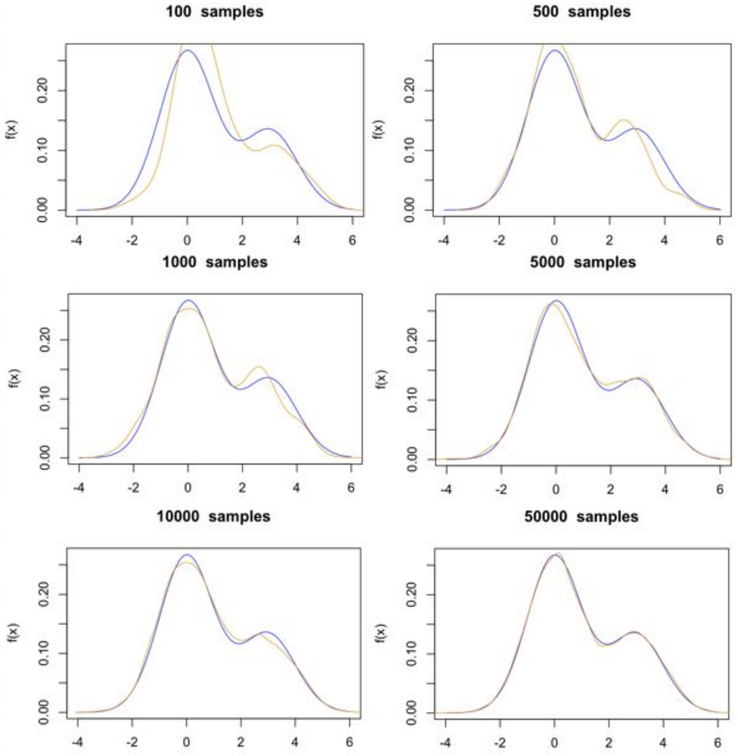
$$\int \pi(x)dx = \int e^{-H(x)/k_B T}dx = Z$$

Recap on Markov Chain Monte Carlo

Construct a Markov Chain X_i stationary and ergodic against ρ.

- Averages over X_i converge to averages over ρ.

$$\bar{f} := \frac{1}{N}\sum_{i=1}^{N} f\left(X_i\right) \to \frac{\int f(x)\rho(x)dx}{\int \rho(x)dx}$$



Recap on Markov Chain Monte Carlo

Metropolis-Hastings:

1. Draw a proposal for the next point

$$Y_{i+1} = Q(Y|X_i)$$

2. Accept the proposal with probability.

$$p_{accept} = \min\left\{1, \frac{\rho(Y_{i+1})Q(Y_{i+1}|X_i)}{\rho(X_i)Q(X_i|Y_{i+1})}\right\}$$

3. If accepted, $X_{i+1} = Y_{i+1}$. Else, $X_{i+1} = X_i$.

(In practice, more likely to a method like Hamiltonian MC, no U-turn, inertial Langevin…)

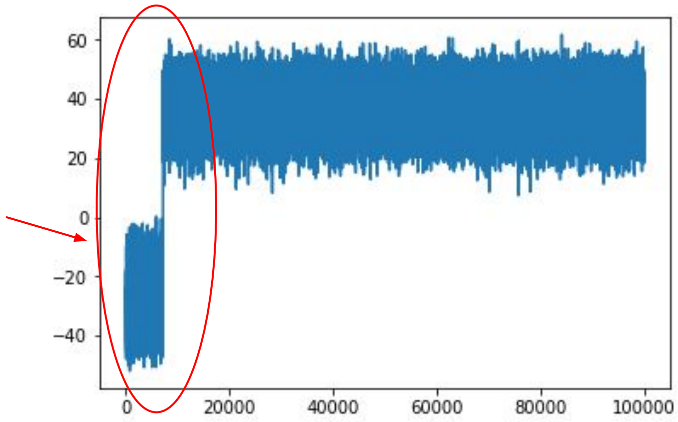
1. How can check if our MCMC is converging?

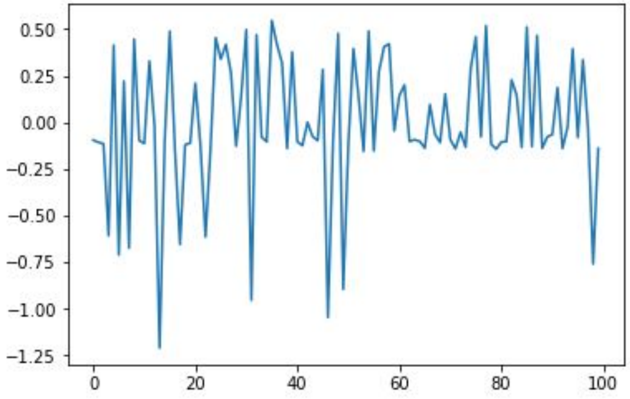
2. How can we improve convergence?

Tools for MCMC Convergence

Visual inspection can tell us how well chains are converged.

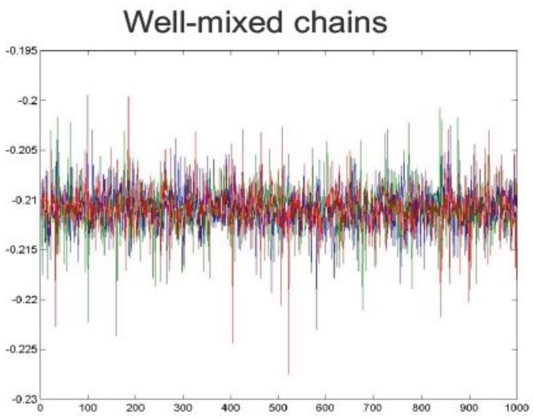
Burn-in / Warmup period

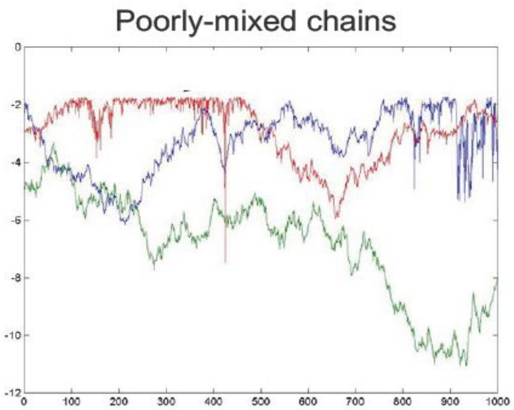




Tools for MCMC Convergence

- Run chains with different initial points.
- $\hat{R}$ statistic compares variance of each chain to variance of the ensemble.
  - If perfectly converged, $\hat{R} = 1$.
  - If ensemble variance >> chain variance, $\hat{R} \gg 1$.
  - A good threshold is $\hat{R} \leq 1.01$





Tools for MCMC Convergence

Can we *quantitatively* assess convergence?

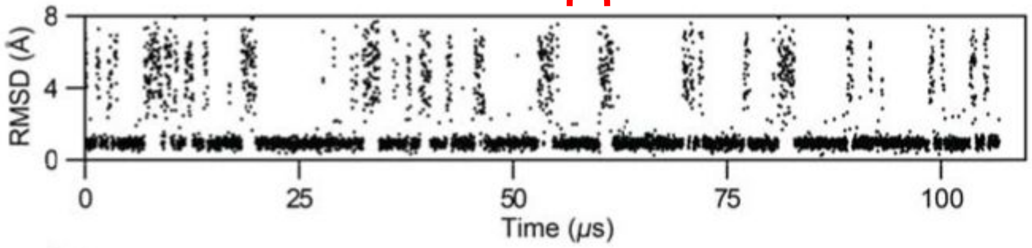
$$\text{var}\left(\bar{f}\right) \approx \frac{\tau\,\text{var}(f)}{N}$$

How long it takes Markov Chain to decorrelate and give a new "independent" sample.

Effective Sample Size (ESS): N/τ

- >100 is a good start

$\tau \approx$



Tools for MCMC Convergence

$$\mathrm{var}\left(\bar{f}\right) \approx \frac{\tau\, \mathrm{var}(f)}{N}$$

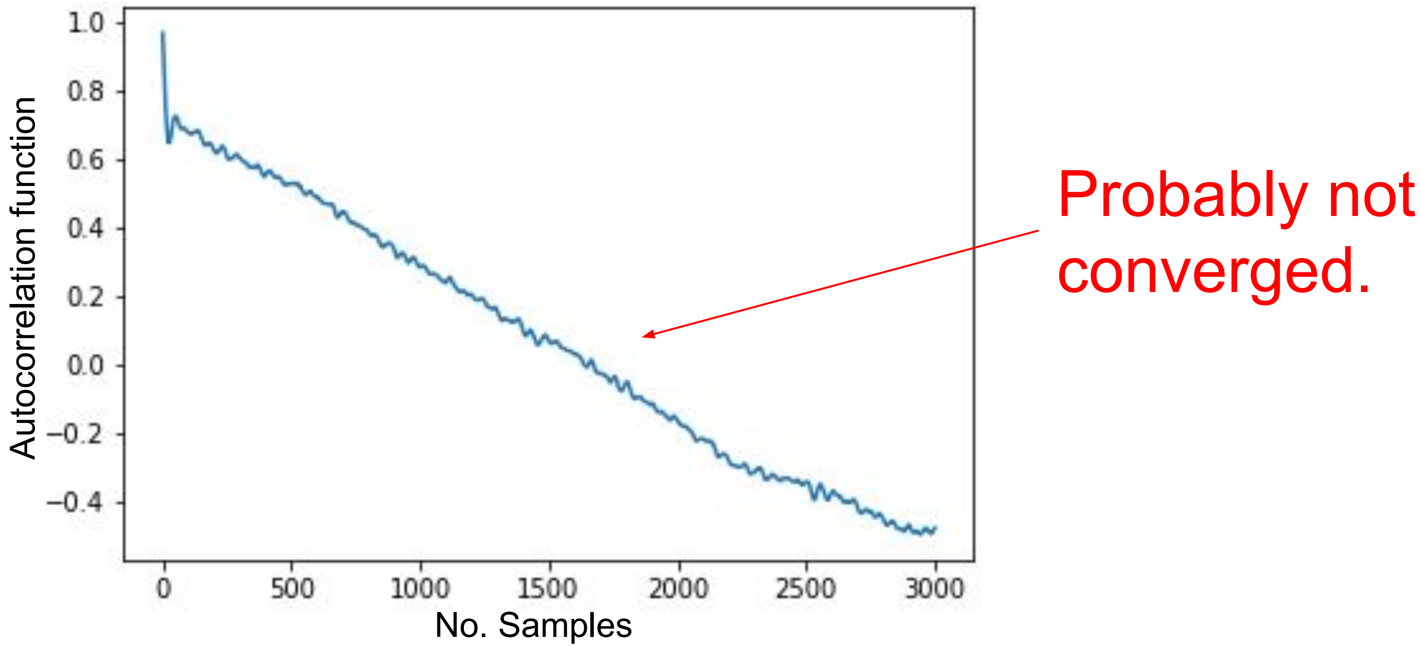
$$\tau = \frac{1}{2} + \frac{1}{2}\sum_{i=1}^{\infty} \frac{\mathbf{E}\left[\left(f(X_i)-\mu_f\right)\left(f(X_0)-\mu_f\right)\right]}{\mathbf{E}\left[\left(f(X_0)-\mu_f\right)\left(f(X_0)-\mu_f\right)\right]}$$

Hard to compute: variance of naive estimator diverges!

- ESS in Stan
- autocorr in emcee

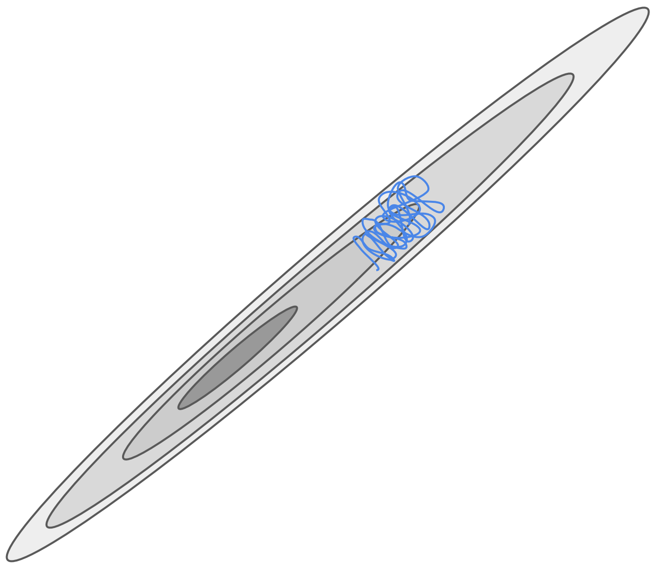
Tools for MCMC Convergence

Look at your autocorrelation functions!

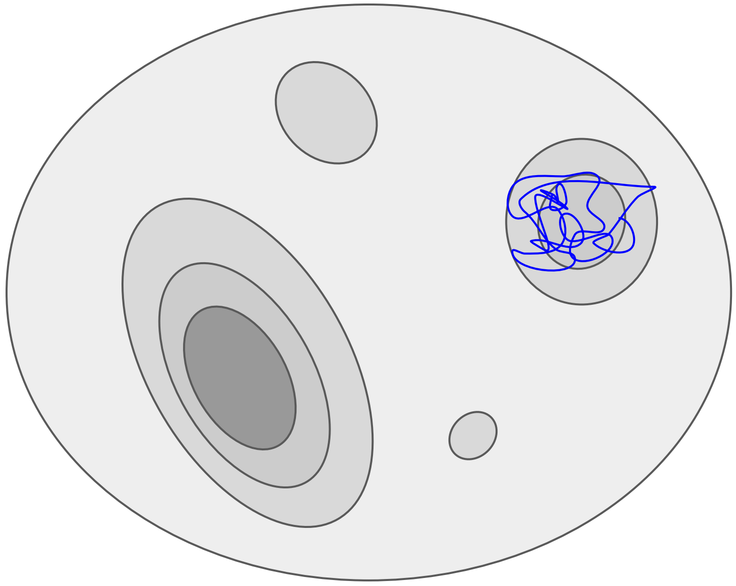


But *why* is my MCMC converging slowly?

Bad Scaling



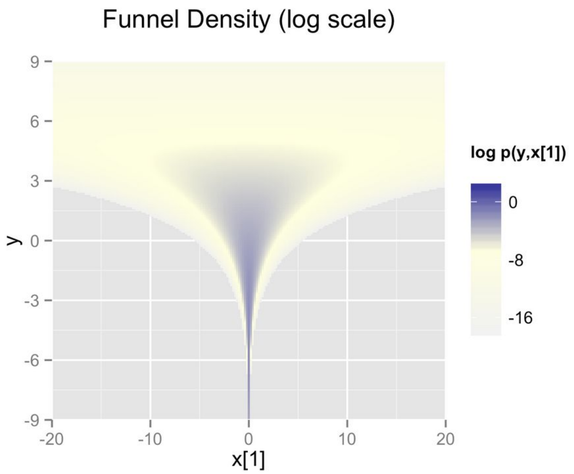
Multimodality



Fixing Bad Scaling

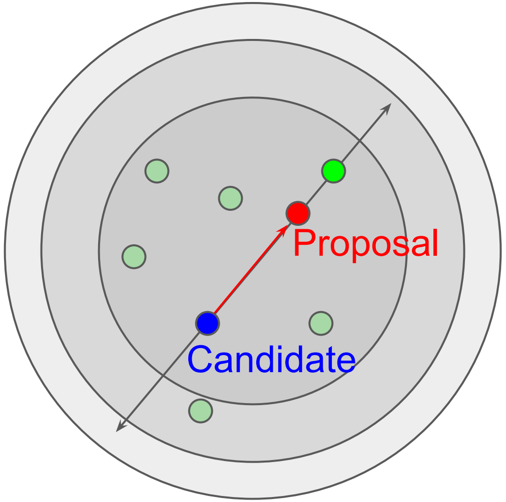
MCMC moves don't match scaling for your problem.

- Reparameterize model
  - Think!
- Scale problem so that Hessian is close to $\mathcal{I}$
  - Explicitly evaluate Hessian?
- Estimate covariance from Warmup runs (automatic in Stan)

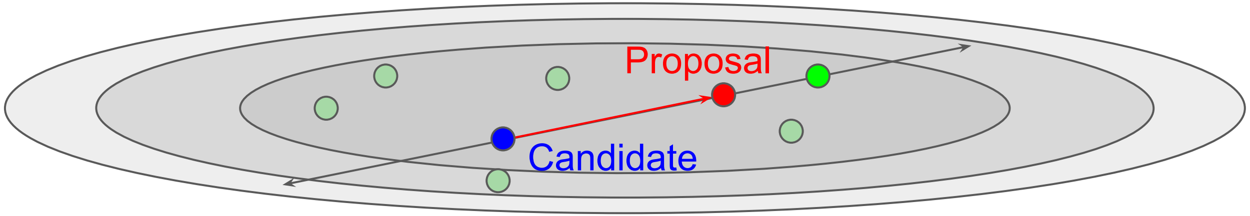


Affine Invariant Sampling

Ensure MCMC chain convergence is unchanged by stretching and translating the parameters.



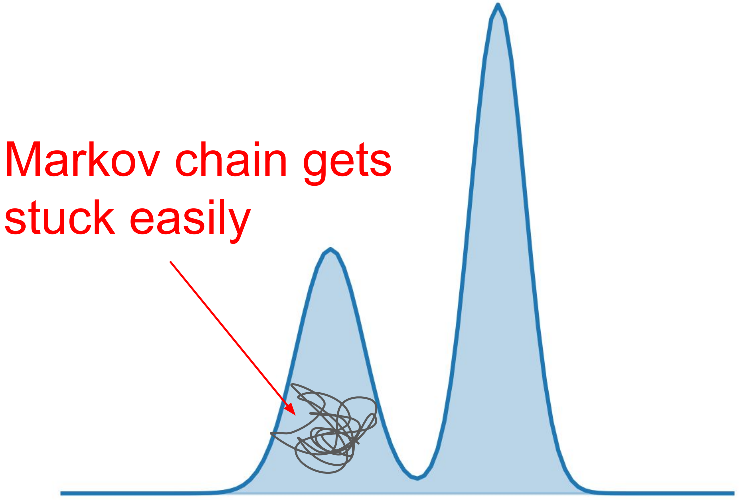
The Affine Invariant Sampler (what emcee is famous for)



Multimodality

With many modes, MCMC must cross low probability regions.

- Hard to solve by changing MCMC proposals
- Change the probability distribution instead.
- Related problem: sampling tail probabilities.



Multimodality (A side note for the Bayesians)

Recall:

Parameters (what gets sampling)

$$\rho(\theta) = p\left(y|\theta\right) p\left(\theta\right)$$

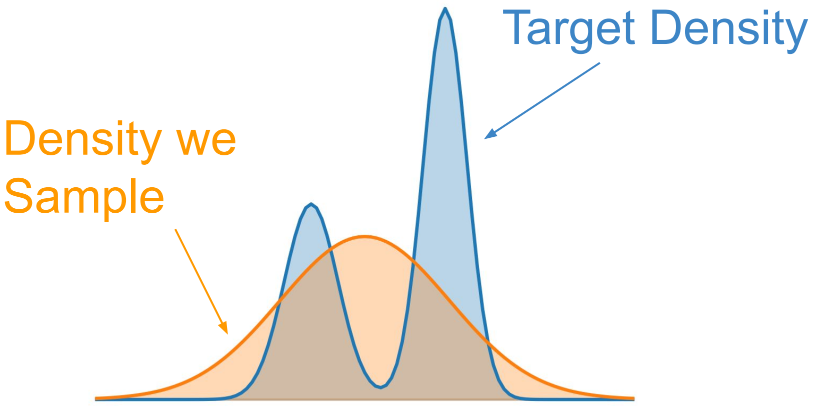
$$\int \pi(\theta)dx = \int_\theta p\left(y|\theta\right) p\left(\theta\right) d\theta = p(y)$$

If the model family contains the right model, we should see one peak for the real parameters.

Multimodality $\xrightarrow{???}$ Model misspecification

Importance Sampling

Sample a different distribution π and reweight.



$$\frac{\int f(x)\rho(x)dx}{\int \rho(x)dx} = \frac{\int f(x)\frac{\rho(x)}{\pi(x)}\pi(x)dx}{\int \frac{\pi(x)}{\pi(x)}\pi(x)dx}$$

$$\approx \frac{\sum_{i=1}^{N} f(X_i)\frac{\rho(X_i)}{\pi(X_i)}}{\sum_{i=1}^{N} \frac{\rho(X_i)}{\pi(X_i)}}$$

...but if you choose badly, variance can be infinite.

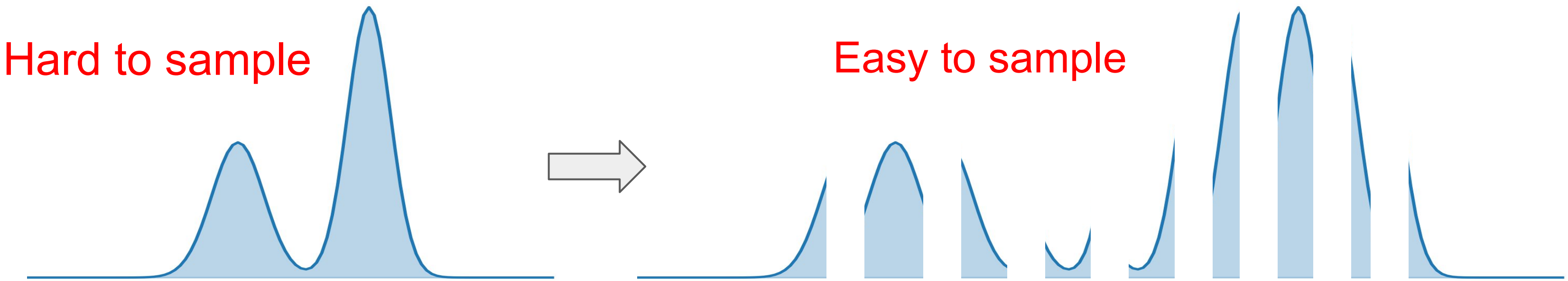
Umbrella Sampling

Why not reweight using many states?

- Introduce biases $\psi_j(x)$ and sample biased distribution

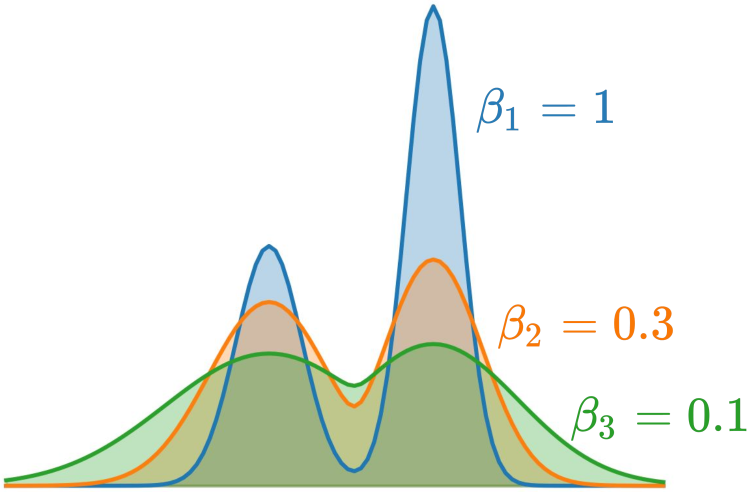
$$\pi_j = \frac{\psi_j(x)\rho(x)dx}{\int \psi_j(x)\rho(x)dx}$$

Determined using overlaps between biased states.



Replica Exchange

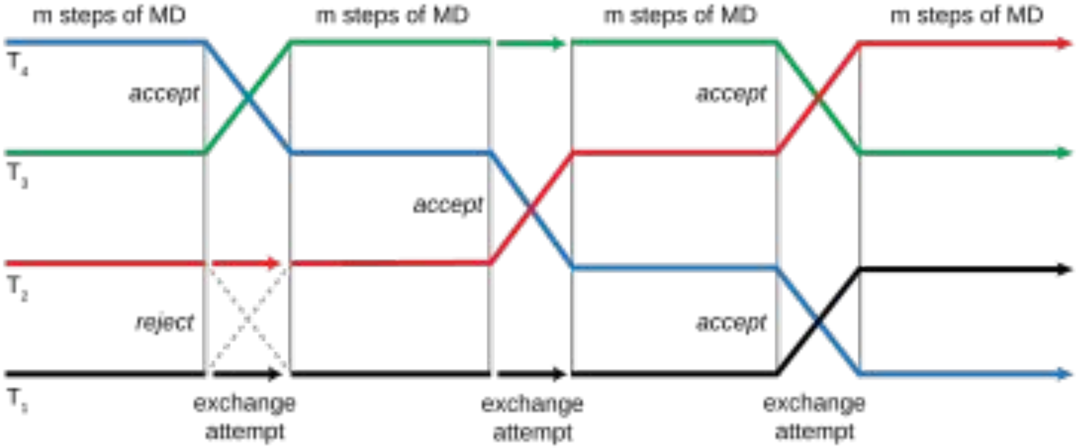
US biases don't have to be spatial: $\pi_j(x) = \rho(x)^{\beta_j}$



But if we sample these distributions, β_1 still won't converge.

Replica Exchange

Idea: Use other probability distributions as route to jump between modes.



- Run one Markov chain in each state π_j.
- Periodically attempt swaps between states.
  - If done correctly, marginal in each state is π_j.
- Should generally be done *along with* US reweighting.

What if I don't know the density?

~~Known exactly~~ unknown

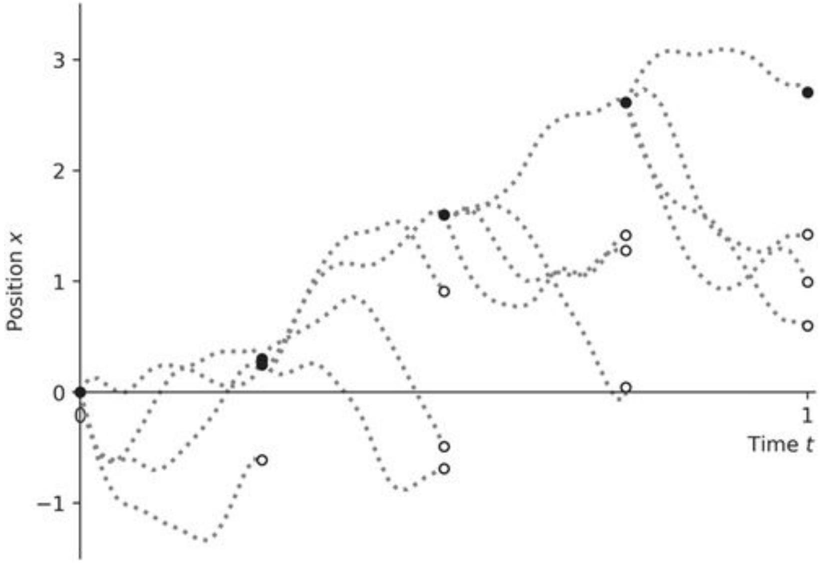
$$\mu_f = \frac{\int f(x)\rho(x)dx}{\int \rho(x)dx}$$

Sometimes distribution is only implicitly defined

- Output of a simulation
- Unknown stationary distribution of a known Markov Chain

One solution: Sequential Monte Carlo

Sequential Monte Carlo



We want the effect of importance sampling, but can't change the MCMC update.

1. Run a swarm of MCMC chains.
2. Split / kill chains
   a. Split chains moving towards good regions.
   b. Kill chains moving towards bad regions.

Conclusions

Bad MCMC convergence: often bad scaling and multimodality.

For bad scaling, try

- Reparameterization
- Calculating the Hessian
- Ensemble Methods

For multimodality, try

- Importance Sampling
- Umbrella Sampling
- Replica Exchange

If density is unknown, try Sequential MCMC.

A highly non-exhaustive bibliography.

Bibliography: Basic MCMC resources.

For Bayesian problems

- Stan Tutorials, at *https://mc-stan.org/users/documentation/tutorials.html*

For Physical Systems

- Frenkel, Daan, and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Vol. 1. Elsevier, 2001.
- Sokal, Alan. "Monte Carlo methods in statistical mechanics: foundations and new algorithms." *Functional integration*. Springer, Boston, MA, 1997. 131-192.

For Both

- Liu, Jun S. *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

Bibliography: Convergence

R-hat statistic

- Initial version: Gelman, Andrew, and Donald B. Rubin. "Inference from iterative simulation using multiple sequences." *Statistical science* 7.4 (1992): 457-472.
- The latest version: Vehtari, Aki, et al. "Rank-normalization, folding, and localization: An improved $\widehat {R} $ for assessing convergence of MCMC." *Bayesian Analysis* (2020).

Autocorrelation Time

- Geyer, Charles J. "Practical markov chain monte carlo." *Statistical science* (1992): 473-483.
- Goodman, J. "Acor, statistical analysis of a time series." (2009). https://www.math.nyu.edu/faculty/goodman/software/acor/index.html

Bibliography: Addressing Bad Scaling

Affine Invariant Sampler

- Goodman, Jonathan, and Jonathan Weare. "Ensemble samplers with affine invariance." *Communications in applied mathematics and computational science* 5.1 (2010): 65-80.
- Foreman-Mackey, Daniel, et al. "emcee: the MCMC hammer." *Publications of the Astronomical Society of the Pacific* 125.925 (2013): 306.

Some Other Approaches

- Calderhead, Ben. *Differential geometric MCMC methods and applications*. Diss. University of Glasgow, 2011.
- Roberts, Gareth O., and Jeffrey S. Rosenthal. "Examples of adaptive MCMC." *Journal of Computational and Graphical Statistics* 18.2 (2009): 349-367.

Bibliography: Multimodality

Importance Sampling

- Kahn, Herman, and Theodore E. Harris. "Estimation of particle transmission by random sampling." *National Bureau of Standards applied mathematics series* 12 (1951): 27-30.

Umbrella Sampling:

- Optimal Recombination procedure in:
    - Vardi, Yehuda. "Empirical distributions in selection bias models." *The Annals of Statistics* (1985): 178-203.
    - Shirts, Michael R., and John D. Chodera. "Statistically optimal analysis of samples from multiple equilibrium states." *The Journal of chemical physics* 129.12 (2008): 124105.
- Recommended Solver (iterative EMUS) in
    - Thiede, Erik H., et al. "Eigenvector method for umbrella sampling enables error analysis." *The Journal of chemical physics* 145.8 (2016): 084115.

Bibliography: Multimodality cont.

Replica Exchange:

- Swendsen, Robert H., and Jian-Sheng Wang. "Replica Monte Carlo simulation of spin-glasses." *Physical review letters* 57.21 (1986): 2607.
- Geyer, Charles J. "Markov chain Monte Carlo maximum likelihood." (1991).
- Earl, David J., and Michael W. Deem. "Parallel tempering: Theory, applications, and new perspectives." *Physical Chemistry Chemical Physics* 7.23 (2005): 3910-3916.

Sequential Monte Carlo

- Huber, Gary A., and Sangtae Kim. "Weighted-ensemble Brownian dynamics simulations for protein association reactions." *Biophysical journal* 70.1 (1996): 97-110.
- See chapter in book by Liu, Jun S.
- Recent overview / theoretical analysis: Webber, Robert J. "Unifying sequential Monte Carlo with resampling matrices." *arXiv preprint arXiv:1903.12583* (2019).