Hamiltonian Monte Carlo & Convegence

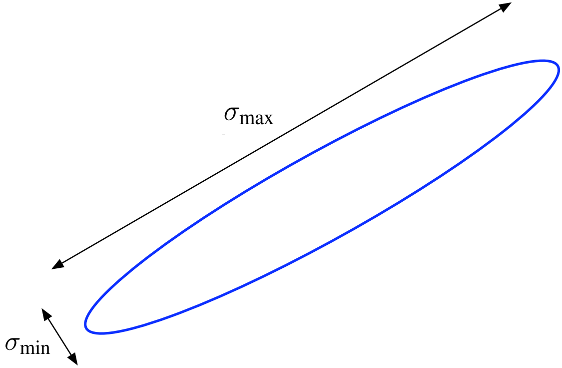
Donald E. Brown

School of Data Science
University of Virginia
Charlottesville, VA 22904

Hamiltonian Approach

- MCMC is inefficient
- Hamiltonian Monte Carlo (HMC)
  - Use gradient information for the proposal step
  - Accept or reject using MH criterion
- Pros and cons of HMC
  - Faster moves to interesting regions
  - Costs more to evaluate
  - Only for continuous distributions

Example



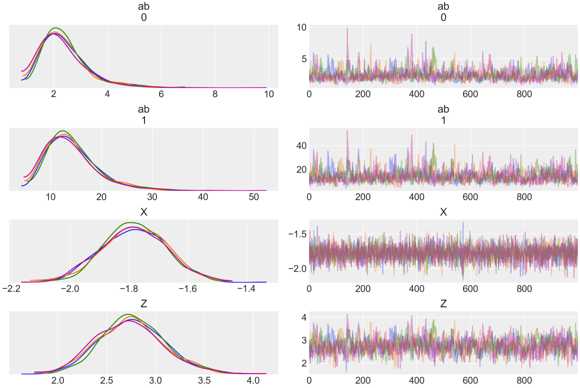
MCMC Convergence

- The Markov chain will eventually converge to the target distribution but when?
- There are no provable guarantees of convergence
- Several simple tests that help us decide if the chain has converged

Graphics and Visualizations

- *Traceplots* plot the value of the draw vs. the iteration number and show how well the chain is sampling or *mixing* the parameter space
  - Show if the chain has gotten stuck in a an area of the sample space
  - Make one for every parameter
- *Density plots* show the skew in the sampling

Chain Diagnostic Plot Examples



Multiple Sequence Diagnostics

- Variance within a chain: W
- Between chain variance for M chains, each of length N

$$B = \frac{N}{M-1}\sum_{j=1}^{M}(\bar{\theta}_j - \bar{\bar{\theta}})^2, \quad \bar{\theta}_j = \frac{1}{N}\sum_{i=1}^{N}\theta_{ij}, \quad \bar{\bar{\theta}} = \frac{1}{M}\sum_{j=1}^{M}\bar{\theta}_j$$

$$\widehat{\text{Var}}(\theta) = (1 - \frac{1}{N})W + \frac{1}{N}B$$

- Scale reduction or shrinkage factor

$$\hat{R} = \left\{\frac{\widehat{\text{Var}}(\theta)}{W}\right\}^{\frac{1}{2}}$$

- Run longer if $\hat{R} > 1.2$

Summary of Sampling Methods

- Highly accurate
- Handles complexity
- Slow - scaling issues
- New sampling methods occur with frequency