Dynamic Hamiltonian Monte Carlo in Stan

- Hamiltonian Monte Carlo
  - use of gradient information and dynamic simulation reduce random walk
- Dynamic HMC
  - adaptive simulation time
- Adaptation of algorithm parameters
  - mass matrix and step size adaptation during warm-up
- Dynamic HMC specific diagnostics

Extra material for dynamic HMC

- Michael Betancourt (2018). Scalable Bayesian Inference with Hamiltonian Monte Carlo https://www.youtube.com/watch?v=jUSZboSq1zg
- Michael Betancourt (2018). A Conceptual Introduction to Hamiltonian Monte Carlo. https://arxiv.org/abs/1701.02434
- http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/
- Cole C. Monnahan, James T. Thorson, and Trevor A. Branch (2016) Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. https://dx.doi.org/10.1111/2041-210X.12681

Demos

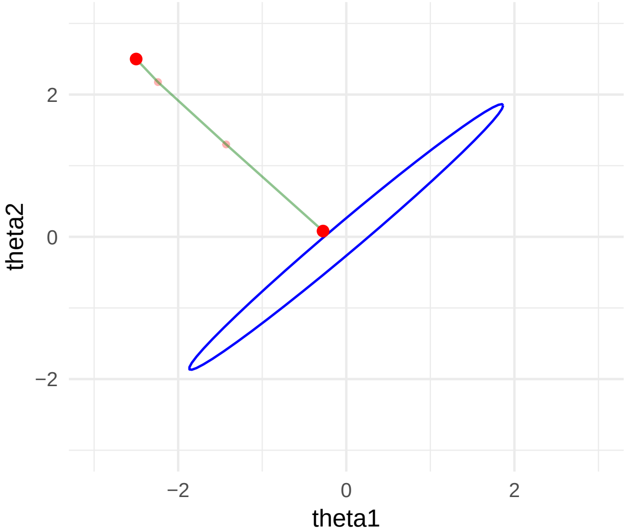
- https://github.com/avehtari/BDA_R_demos/tree/master/demos_ch12
  - demos_ch12/demo12_1.R
- http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/

Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables

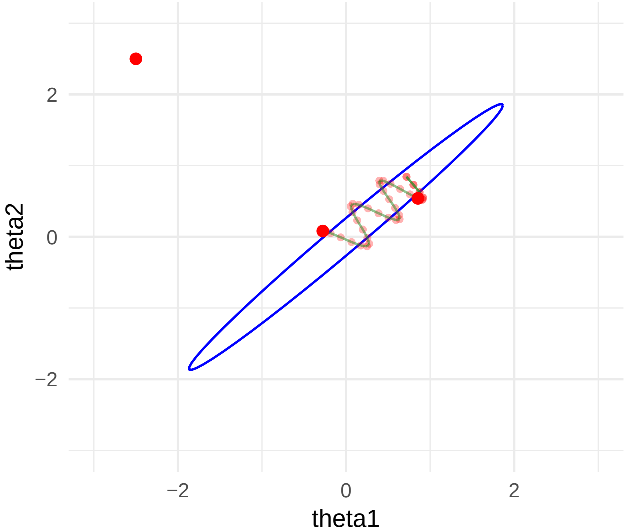
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



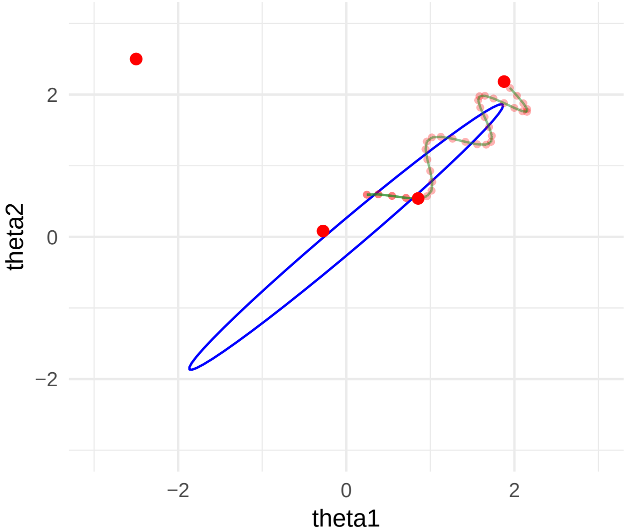
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



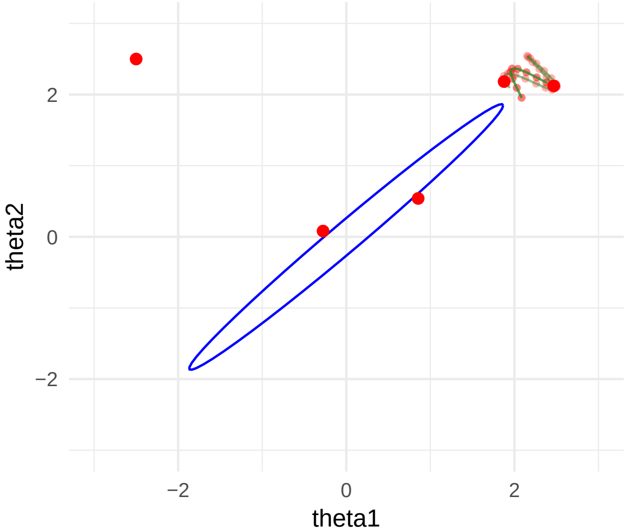
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



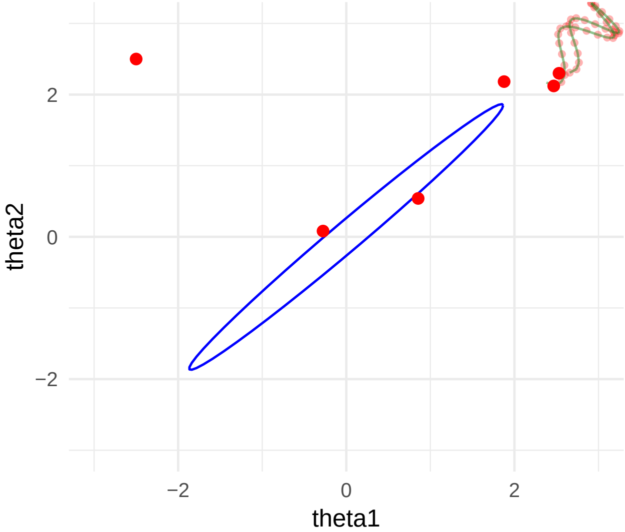
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



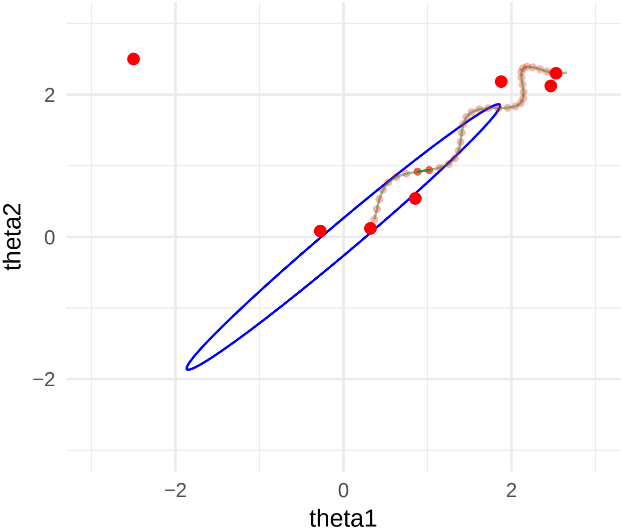
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



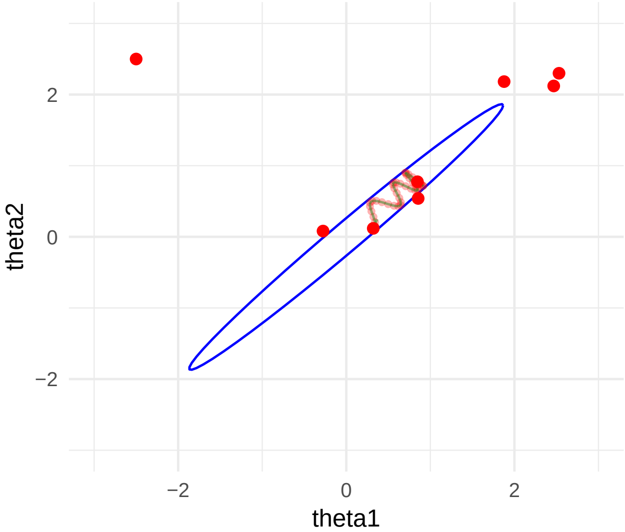
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



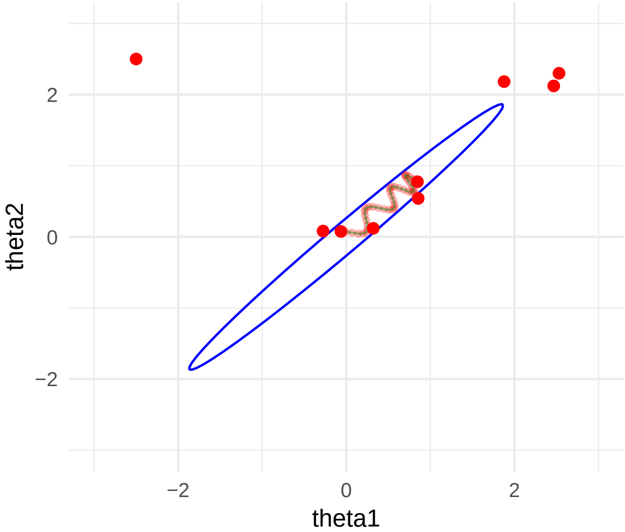
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



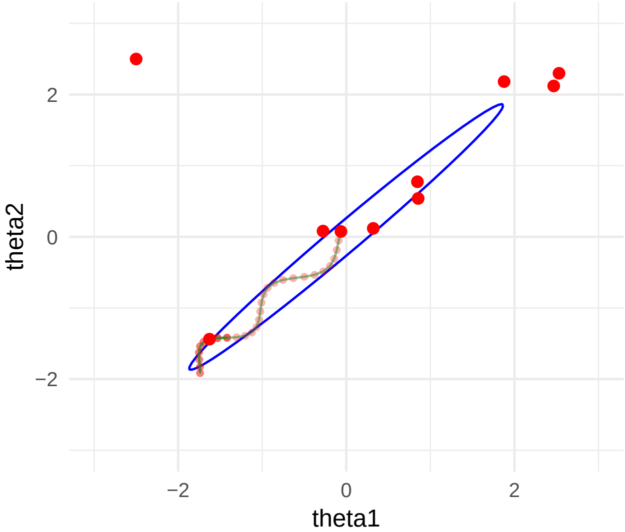
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



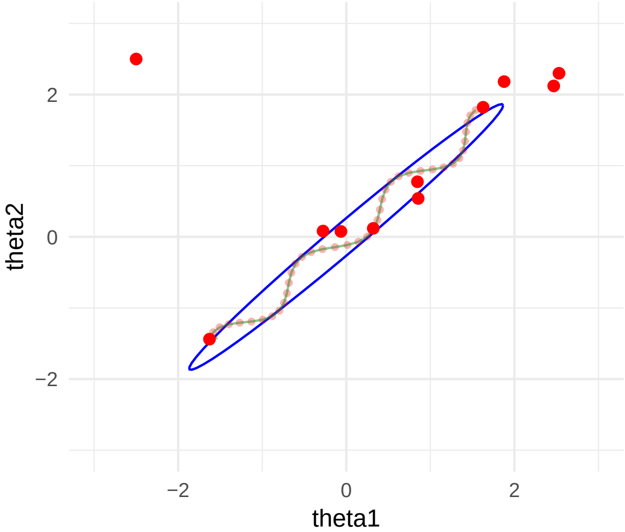
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



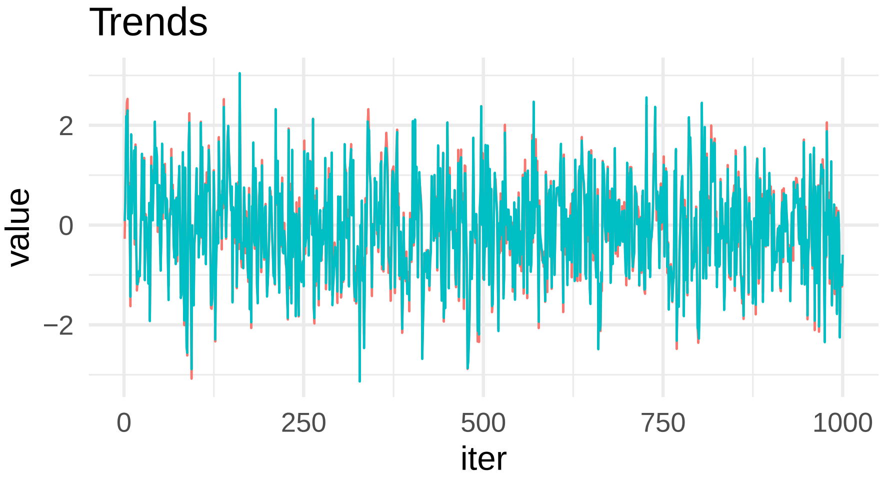
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



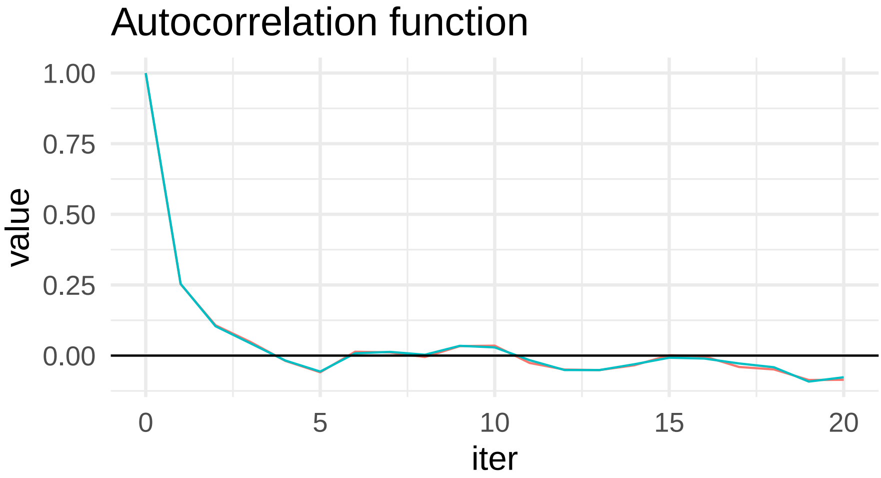
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



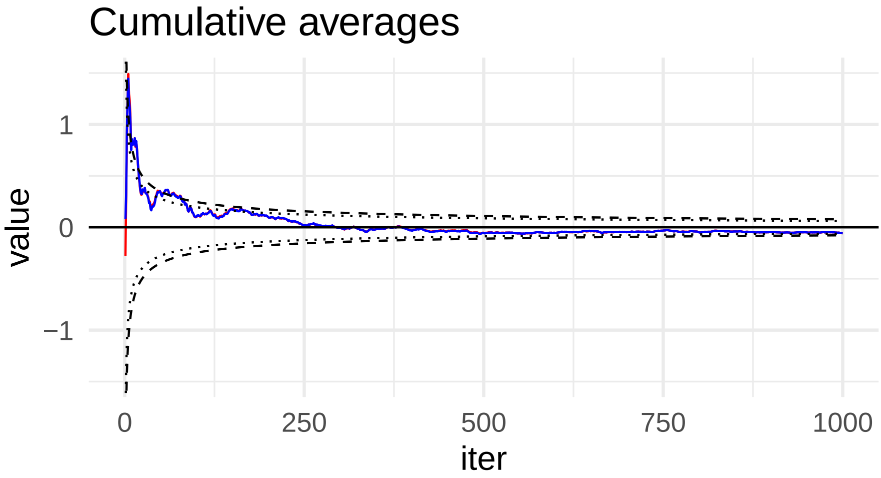
Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables



Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Augments parameter space with momentum variables
- Simulation of Hamiltonian dynamics reduces random walk
  - http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/

Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Alternating dynamic simulation and sampling of the energy level

Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Alternating dynamic simulation and sampling of the energy level
- Parameters: step size, number of steps in each chain

Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Alternating dynamic simulation and sampling of the energy level
- Parameters: step size, number of steps in each chain
- No U-Turn Sampling (NUTS) and dynamic HMC
  - adaptively selects number of steps to improve robustness and efficiency
  - dynamic HMC refers to dynamic trajectory length
  - to keep reversibility of Markov chain, need to simulate in two directions
  - http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/

Hamiltonian Monte Carlo

- Uses gradient information for more efficient sampling
- Alternating dynamic simulation and sampling of the energy level
- Parameters: step size, number of steps in each chain
- No U-Turn Sampling (NUTS) and dynamic HMC
  - adaptively selects number of steps to improve robustness and efficiency
  - dynamic HMC refers to dynamic trajectory length
  - to keep reversibility of Markov chain, need to simulate in two directions
  - http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/
- Dynamic simulation is discretized
  - small step size gives accurate simulation, but requires more log density evaluations
  - large step size reduces computation, but increases simulation error which needs to be taken into account in the Markov chain

Adaptive dynamic HMC in Stan

- Dynamic HMC using growing tree to increase simulation trajectory until no-U-turn criterion stopping
  - max treedepth to keep computation in control
  - pick a draw along the trajectory with probabilities adjusted to take into account the error in the discretized dynamic simulation

Adaptive dynamic HMC in Stan

- Dynamic HMC using growing tree to increase simulation trajectory until no-U-turn criterion stopping
  - max treedepth to keep computation in control
  - pick a draw along the trajectory with probabilities adjusted to take into account the error in the discretized dynamic simulation
- Mass matrix and step size adaptation in Stan
  - mass matrix refers to having different scaling for different parameters and optionally also rotation to reduce correlations
  - mass matrix and step size adjustment and are estimated during initial adaptation phase
  - step size is adjusted to be as big as possible while keeping discretization error in control

Adaptive dynamic HMC in Stan

- Dynamic HMC using growing tree to increase simulation trajectory until no-U-turn criterion stopping
  - max treedepth to keep computation in control
  - pick a draw along the trajectory with probabilities adjusted to take into account the error in the discretized dynamic simulation
- Mass matrix and step size adaptation in Stan
  - mass matrix refers to having different scaling for different parameters and optionally also rotation to reduce correlations
  - mass matrix and step size adjustment and are estimated during initial adaptation phase
  - step size is adjusted to be as big as possible while keeping discretization error in control
- After adaptation the algorithm parameters are fixed and some further iterations included in the warmup

Adaptive dynamic HMC in Stan

- Dynamic HMC using growing tree to increase simulation trajectory until no-U-turn criterion stopping
  - max treedepth to keep computation in control
  - pick a draw along the trajectory with probabilities adjusted to take into account the error in the discretized dynamic simulation
- Mass matrix and step size adaptation in Stan
  - mass matrix refers to having different scaling for different parameters and optionally also rotation to reduce correlations
  - mass matrix and step size adjustment and are estimated during initial adaptation phase
  - step size is adjusted to be as big as possible while keeping discretization error in control
- After adaptation the algorithm parameters are fixed and some further iterations included in the warmup
- After warmup store iterations for inference

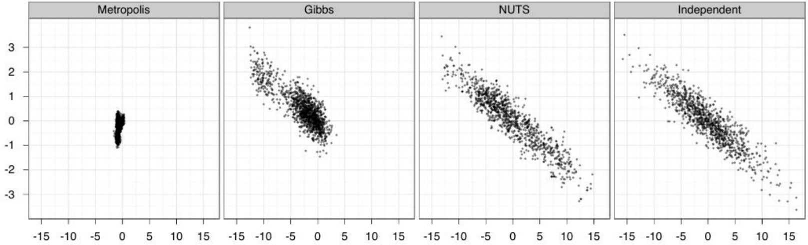
Adaptive dynamic HMC in Stan

- Dynamic HMC using growing tree to increase simulation trajectory until no-U-turn criterion stopping
  - max treedepth to keep computation in control
  - pick a draw along the trajectory with probabilities adjusted to take into account the error in the discretized dynamic simulation
- Mass matrix and step size adaptation in Stan
  - mass matrix refers to having different scaling for different parameters and optionally also rotation to reduce correlations
  - mass matrix and step size adjustment and are estimated during initial adaptation phase
  - step size is adjusted to be as big as possible while keeping discretization error in control
- After adaptation the algorithm parameters are fixed and some further iterations included in the warmup
- After warmup store iterations for inference
- See more details in Stan reference manual

Dynamic HMC

Comparison of algorithms on **highly correlated** 250-dimensional Gaussian distribution

- Do **1,000,000** draws with both Random Walk Metropolis and Gibbs, thinning by 1000
- Do **1,000** draws using Stan's NUTS algorithm (no thinning)
- Do 1,000 independent draws (we can do this for multivariate normal)



Source: Jonah Gabry

Max tree depth diagnostic

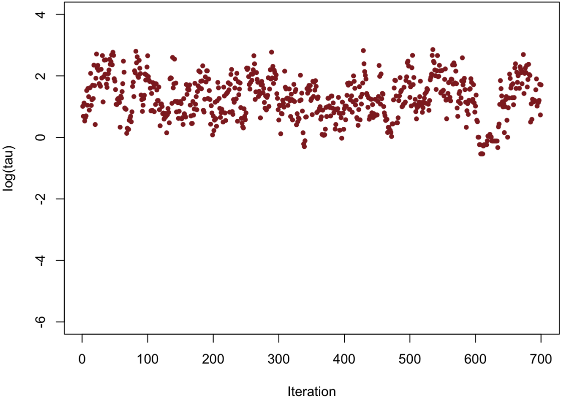
- Dynamic HMC specific diagnostic
- Indicates inefficiency in sampling leading to higher autocorrelations and lower n_{eff}
- Different parameterizations matter

Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html

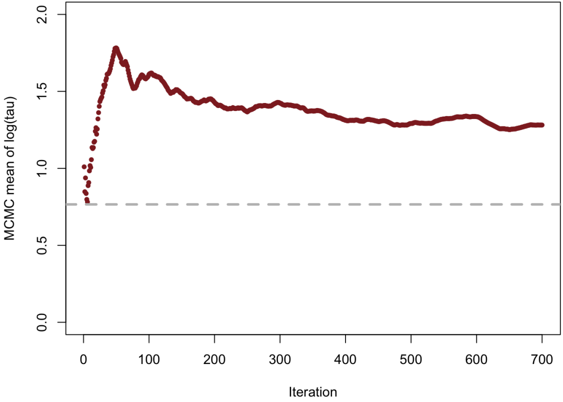
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



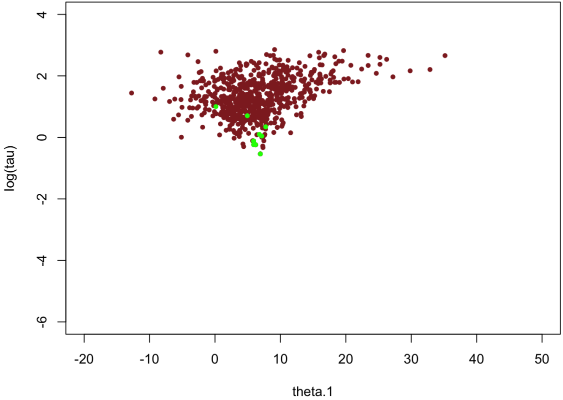
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



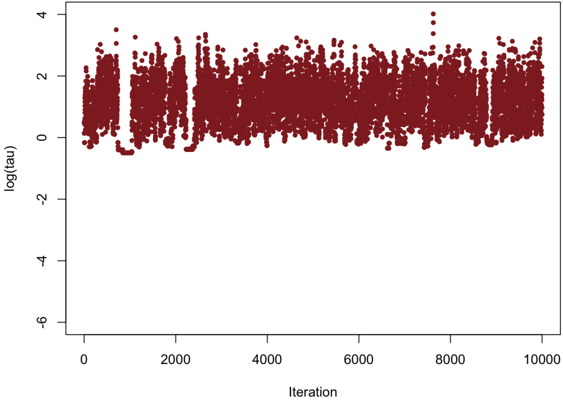
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



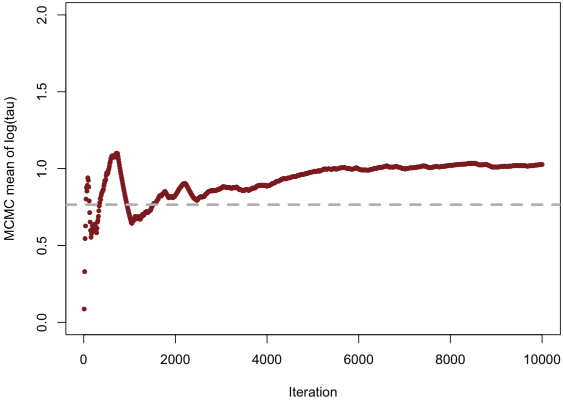
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



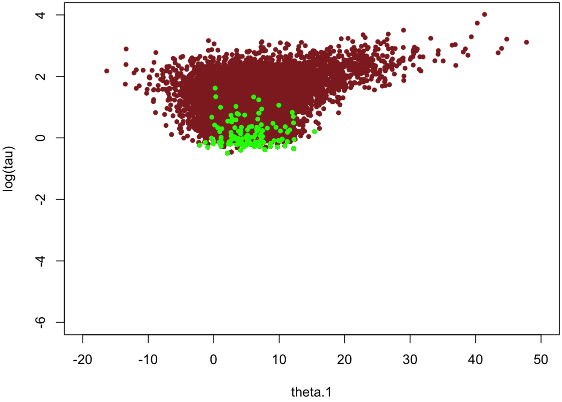
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



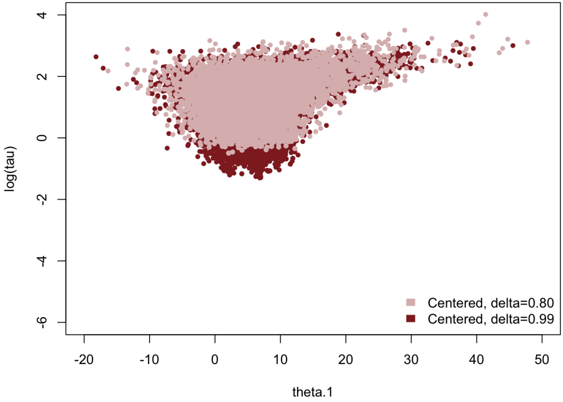
Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html



Divergences

- HMC specific: Indicates that Hamiltonian dynamic simulation has problems going to narrow places
  - indicates possibility of biased estimates
- Different parameterizations matter
- http://mc-stan.org/users/documentation/case-studies/divergences_and_bias.html