Langevin Monte Carlo

Final Project for Intro to Bayesian Computing

Anthony Bugatto

Monte Carlo Methods

- Monte Carlo methods are a general framework that rely on sampling statistical distributions to obtain numerical results
- Have wide usage in physics, engineering, and natural science
- Developed during the Manhattan Project with large contributions from:
    - John Von Neumann (middle)
    - Enrico Fermi (top)
    - Nicolas Metropolis (bottom)

Figure 1: Pioneers of Monte Carlo

Monte Carlo Markov Chain - Metropolis Hastings

- Assumes a set of states with a transition probability from state to state
  - Eventually converges to a stationary distribution
- Can accept or reject samples based on the probability of the target distribution
- Pros:
  - Works well in high dimensions
  - Can sample arbitrary distributions effectively
- Cons:
  - Has difficulty with disjoint multimodal distributions
  - Can take a long time to evenly sampling the distribution

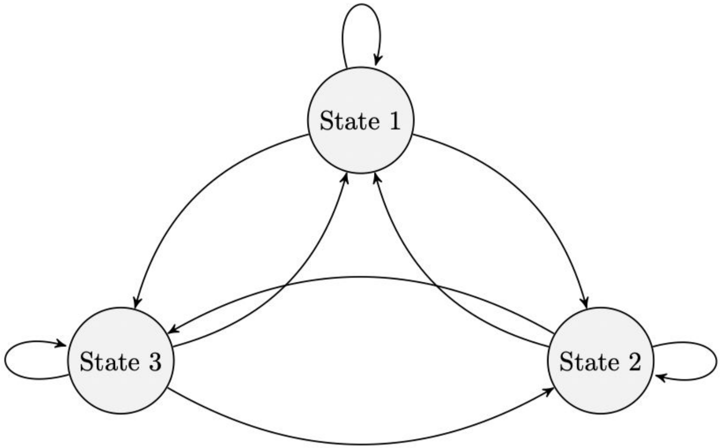


Figure 2: Markov Chain

$$A(x', x) = \min\left(1, \frac{P(x')}{P(x)} \frac{g(x \mid x')}{g(x' \mid x)}\right)$$

Figure 3: Acceptance Probability

Langevin Monte Carlo

- Uses brownian motion and the target distribution gradient to sample across the target more efficiently
- Metropolis Adjusted Langevin Algorithm (MALA) combines langevin dynamics with metropolis hastings accept-reject
  - Doesn’t always converge

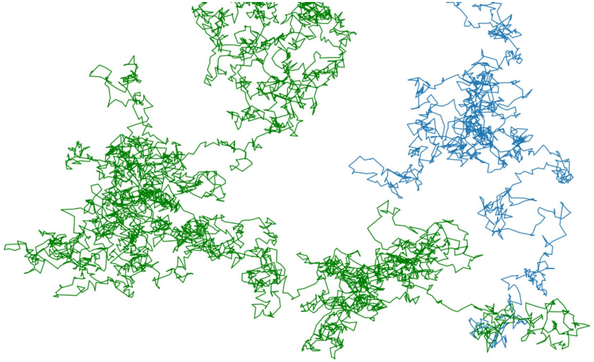


Figure 5: Langevin process motion

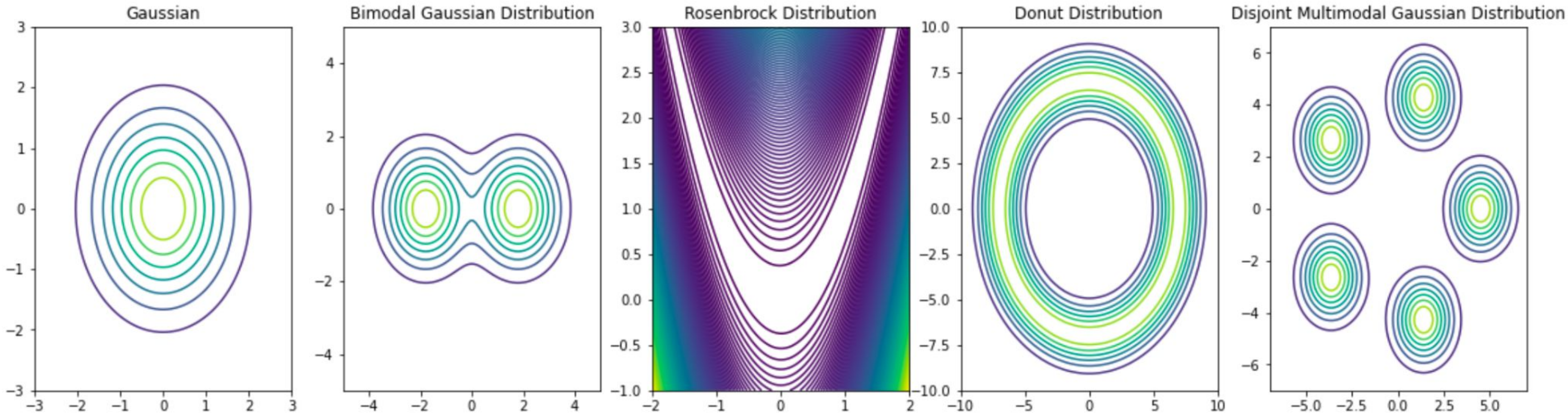
$$\tilde{X}_{k+1} := X_k + \tau \nabla \log \pi(X_k) + \sqrt{2\tau}\xi_k$$

Figure 4: Discrete langevin process transition

$$\alpha := \min\left\{1, \frac{\pi(\tilde{X}_{k+1})q(X_k \mid \tilde{X}_{k+1})}{\pi(X_k)q(\tilde{X}_{k+1} \mid X_k)}\right\}$$

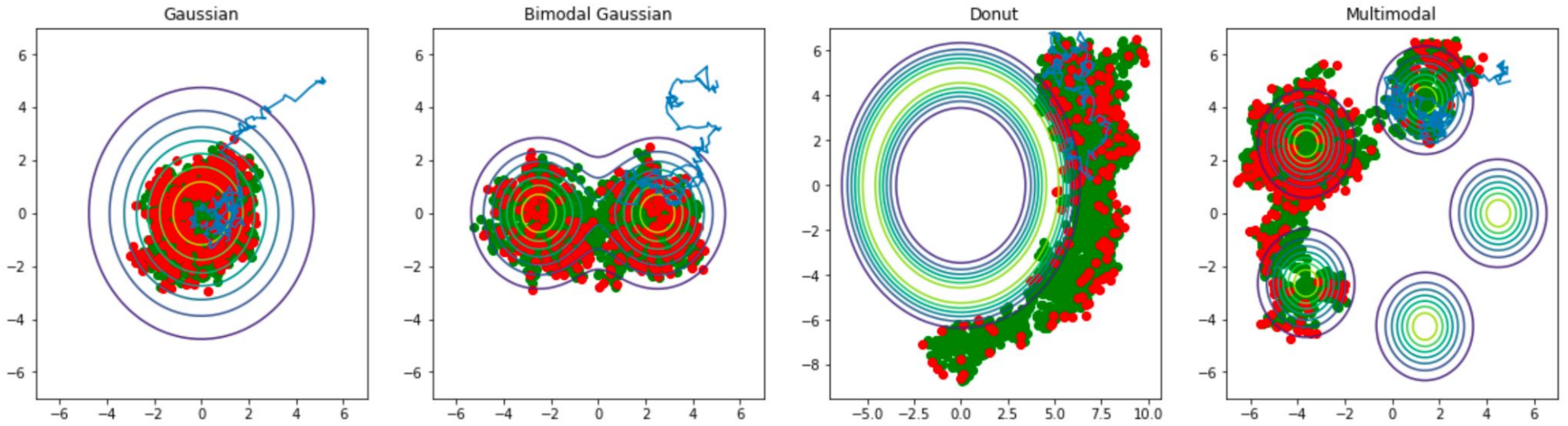
Figure 6: MALA acceptance probability

Test Distributions



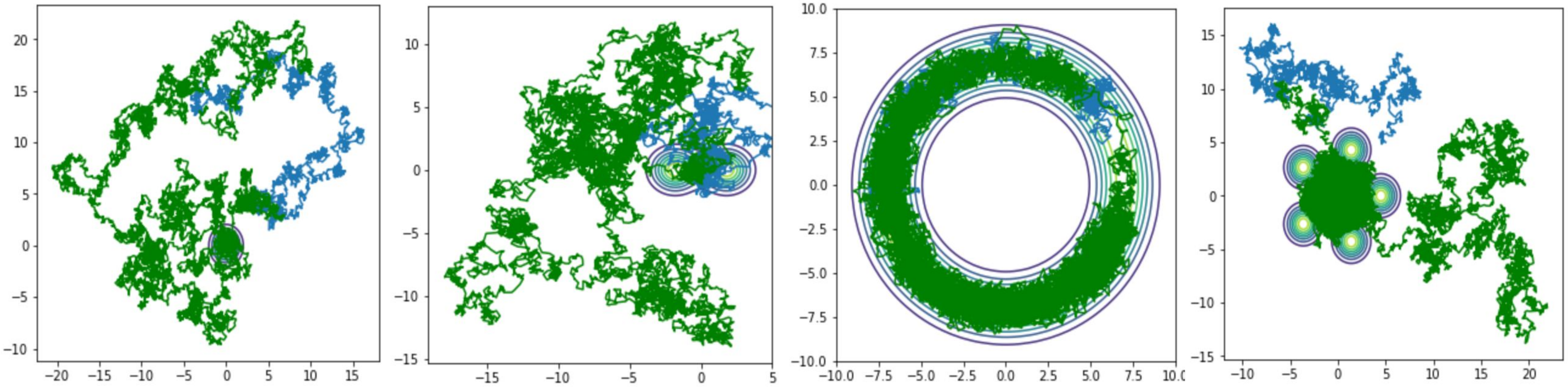
1. **Gaussian Distribution:** Easy to sample using monte carlo and even analytical methods
2. **Bimodal Gaussian Distribution:** Also possible to sample with analytical methods but requires monte carlo to find multiple modes
3. **Rosenbrock Distribution:** Difficult for monte carlo algorithms to evenly sample in reasonable time with low autocorrelation
4. **Donut Distribution:** Difficult for single step gradient based monte carlo algorithms to evenly sample
5. **Disjoint Multimodal Gaussian Distribution:** Difficult for monte carlo algorithms to find each mode without parameter tuning

Results: Metropolis Hastings



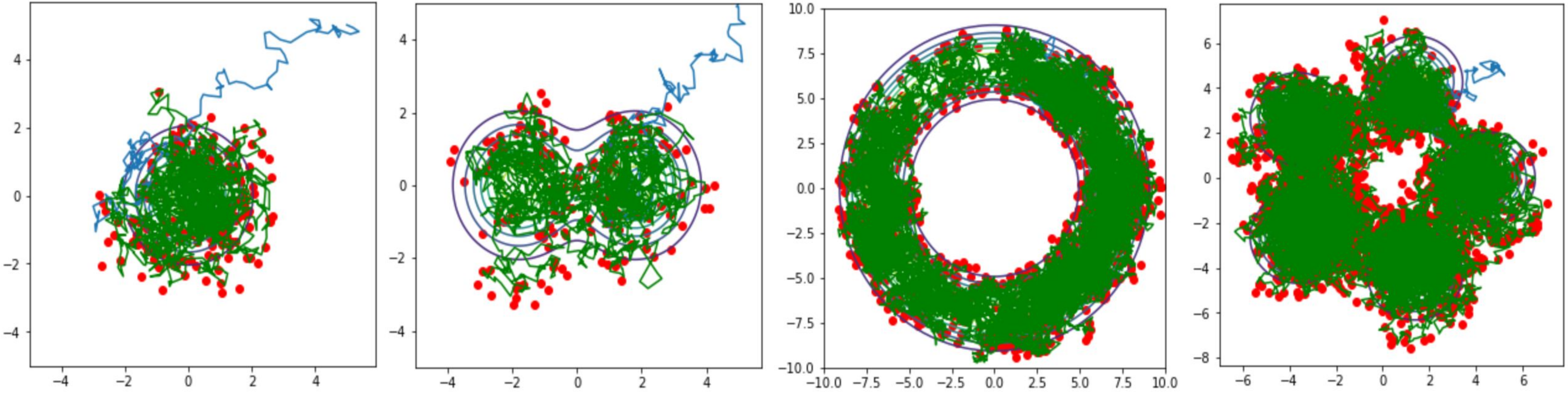
- Burns in in less than 1000 samples
- Generally covers simple distribution
- Covers uncommon very inaccurately given enough samples (10000 for the donut plot)
- Samples multimodal distributions very inaccurately given enough samples (20000 for the right plot)
- Roughly 89% efficiency in sample acceptance

Results: Langevin Monte Carlo



- Often never burns in (all use 20000 samples)
- Generally doesn’t stay in the distribution if it finds it
- Does a good job sampling from uncommonly shaped distributions with the right brownian motion parameters and enough samples

Results: Metropolis Adjusted Langevin Algorithm



- Burns in quickly (<200 samples)
- Needs fewer samples to cover the distribution
- Samples uncommon shapes well given enough samples (10000 for the donut plot)
- Samples multimodal distributions well given enough samples (20000 for the right plot)
- Roughly 91% efficiency in sample acceptance

Conclusions

- Metropolis accept-reject sampling significantly increases efficiency in monte carlo methods
- Langevin monte carlo performs significantly worse than M-H and has trouble converging
- Metropolis adjusted langevin is better at finding multimodal distributions with about equal efficiency as M-H