NTNU | Norwegian University of Science and Technology

THE HASTINGS ALGORITHM AT FIFTY

Paper discussion

Ola Rasmussen & Khanh Truong

January 24, 2024

Contents

Summary

- 1970, W. K. Hastings, Markov chain algorithms, sampling
- Stationary distribution is target distribution
- Hastings improved Metropolis, allowed asymmetry
- Bayesian posterior distributions

1. INTRODUCTION

- Study, characteristics, probability distribution $f(\cdot)$
- If simple, analytically
- If complicated, numerical integration, issues, accuracy, stability, and scalability to higher dimensions
- Solution, Monte Carlo algorithms, estimate features, samples
- Example:
  - Estimate mean
- Solution:
  - With samples $x_t \sim f$, we can estimate the mean of $f(x)$ as $\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} x_t$

1. INTRODUCTION

- Key challenge, efficiently generate samples
- Univariate case, many, inverse cumulative distribution function algorithm is popular
- Arbitrary multivariate distributions, challenging
- Rejection sampling attempts to solve this problem
- Problem, how to select good $g(x)$ that is easy to sample from

1. INTRODUCTION

- Markov chain Monte Carlo algorithms
- Markov chain $\{x_t\}_{t=1}^{T}$, transition kernel $K\left(x_t \mid x_{t-1}\right)$
- Samples $\{x_t\}$, converge, stationary distribution is target distribution
- Burn in, isn't stationary at the beginning
- Particularly popular in Bayesian inference

1. INTRODUCTION

- Metropolis, 1953, built on rejection sampling
- Samples candidate $\tilde{x}$ from proposal density (symmetric)
- Set $x_t = \tilde{x}$ with probability $\alpha\left(\tilde{x} \mid x_{t-1}\right) = min\left\{1, \frac{f(\tilde{x})}{f(x_{t-1})}\right\}$ and $x_t = x_{t-1}$ otherwise
- Big problem, only symmetric distributions

1. INTRODUCTION

- Hastings, 1970
- Improved, asymmetry
- Set $x_t = \tilde{x}$ with probability $\alpha\left(\tilde{x} \mid x_{t-1}\right) = min\left\{1, \frac{f(\tilde{x})}{f(x_{t-1})} \cdot \frac{g(x_{t-1}|\tilde{x})}{g(\tilde{x}|x_{t-1})}\right\}$ and $x_t = x_{t-1}$ otherwise
- Most popular MCMC algorithm
- Most common approach to modern Bayesian computation

1. INTRODUCTION

- Top 10 most important algorithms of the 20th century
- Hydrogen bomb, "mathematical analyzer, numerical integrator, and computer"
- Didn’t mention Bayesian statistics, but is most prominent today
- Made Bayesian statistics feasible

2.1. Overview

- How to choose a good proposal having high computational efficiency?
  (i) Computational cost per iteration of the sampler
  (ii) Mixing rate of the Markov chain $\{x_t\}$
- (i), dependent, cost of sampling, calculating acceptance probability
- (ii), samples are not independent, x_t and $x_{t+\Delta}$ are correlated
- Slow mixing, correlation between x_t and $x_{t+\Delta}$ decreases slowly, samples contribute less information
- Effective sample size

2. Extensions

- Gibbs
- Metropolis-within-Gibbs
- Blocking
- Adaptive algorithms
- Gradient-based algorithms
  - Metropolis-adjusted Langevin
  - Hamiltonian Monte Carlo

3. CHALLENGING APPLICATIONS

Other Pdf

4. EMERGING AREAS & 5. OPEN ARES AND ONGOING CHALLENGES

Other Pdf

Questions

1. Adaptive algorithms
2. What are some advantages of gradient-based algorithms compared to other MCMC methods?
3. What is the best proposal distribution? Should we choose the one that is similar to prior distribution?