Adaptive Markov Chain Monte Carlo Methods

Stefan Junghans
Mitja Richter

TU Berlin

25.07.2014

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Markov Chain Monte Carlo

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Why adaptive?

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Criteria

acceptance rate

- $\alpha = \frac{\#accepted}{\#samples}$
- 0.44 for univariate distributions
- 0.234 for $d \geq 5$ dimensions

suboptimality

- $b = d \cdot \frac{\sum_{i=1}^{d} \lambda_i^{-2}}{(\sum_{i=1}^{d} \lambda_i^{-1})^2}$
- λ_i are eigenvalues of $\Sigma_p^{0.5}\Sigma^{-0.5}$
- Σ_p : proposal covariance matrix
- Σ_p : target covariance matrix
- usually $b > 1$, optimum at $b = 1$

Criteria

autocorrelation time

- $ACT = 1 + 2\sum_{i=1}^{\infty} autocorr(i)$
- $autocorr(k) = 1/((n-k)\cdot v)\cdot\sum_{t=1}^{n-k}(x^t-\mu)(x^{t+k}-\mu)$

average squared jump distance

- $ASJD = \sum_{k=1}^{\infty}\sqrt{\sum_{i=0}^{d}(x_{(i)}^{k}-x_{(i)}^{k+1})^2}$

Outline

1 Introduction
- Markov Chain Monte Carlo
- Why adaptive?
- Criteria

2 Methods
- Metropolis Hastings
- Adaptive Metropolis Hastings
- Adaptive Metropolis Witin Gibbs
- Regional Metropolis Hastings
- Results

3 Summary

Metropolis Hastings

Why do we need it?

- target distribution $f(x)$ that we cannot directly sample from
- $X \sim c \cdot f(x)$, where c is unknown

Metropolis Hastings

Why do we need it?

- target distribution $f(x)$ that we cannot directly sample from
- $X \sim c \cdot f(x)$, where c is unknown

Ingredients

- $f(x)$: target density
- x_0 : start point (random)
- $q(x|y)$: proposal density, often $N(\mu, \Sigma)$

  ($\mu = y \Rightarrow$ "Random Walk Metropolis Hastings")

Metropolis Hastings

Steps

1. current sample : x'
2. sample $x^* \sim q(x|x')$
3. calculate acceptance α:

$$\alpha = min\left(1, \frac{f(x^*)}{f(x')} \cdot \frac{q(x'|x^*)}{q(x^*|x')}\right)$$

4. accept proposal x^* with probability α
5. if accepted : $x' = x^*$
6. repeat

Metropolis Hastings

Example . . .

Outline

1 Introduction
- Markov Chain Monte Carlo
- Why adaptive?
- Criteria

2 Methods
- Metropolis Hastings
- Adaptive Metropolis Hastings
- Adaptive Metropolis Witin Gibbs
- Regional Metropolis Hastings
- Results

3 Summary

Adaptive Metropolis Hastings

Why do we need it?

- If Metropolis Hastings has bad settings, the acceptance rate will be inferior!
- high-dimensional target distributions
- correlated dimensions with different variances

Adaptive Metropolis Hastings

Ingredients

- $f(x)$: target density
- x_0 : start point (random)
- $q(x|y) \sim N(y, 0.1 \cdot I_d/d)$: proposal density

- $q'(x|y) \sim N(y, 2.38^2 \cdot \Sigma_n/d)$: “adapting" proposal density
  - 2.38^2 delivers the best results in certain environments
  - Σ_n is the estimated covariance matrix of the current samples
- β : adaption parameter, e.g. $\beta = 0.05$

Adaptive Metropolis Hastings

Steps

1. current sample : x'
2. if $n \leq 2d$: sample $x^* \sim q(x|x')$
3. if $n > 2d$: sample $x^* \sim (1-\beta) \cdot q'(x|x') + \beta \cdot q(x|x')$
4. calculate acceptance α:

$$\alpha = min\left(1, \frac{f(x^*)}{f(x')}\right)$$

5. accept proposal x^* with probability α
6. if accepted: $x' = x^*$
7. repeat

Adaptive Metropolis Hastings

Example . . .

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Adaptive Metropolis Witin Gibbs

Why do we need it?

- high-dimensional target distributions
- dimensions with different variances

Adaptive Metropolis Witin Gibbs

Ingredients

- $f(x)$: target density
- $x_0 \sim N(0_d, I_d)$: start point
- $q(x|y)$: proposal density
  - ith dimension is $N(x^n_{(i)}, 10^{ls_i})$-distributed
  - $ls_i = 0$ initially
- α_i : individual acceptance rate for every dimension i
- $\delta(n) = min(0.01, n^{-0.5})$: adaption increment
- b : batch size, e.g. $b = 50$

Adaptive Metropolis Witin Gibbs

Steps (1/2)

1. current sample : x'
2. choose dimension i (one after another or randomly)
3. $x^* = x'$
4. $x^*_{(i)} = x^*_{(i)} + N(0, 10^{ls_i})$
5. calculate acceptance α:

$$\alpha = min\left(1, \frac{f(x^*)}{f(x')}\right)$$

6. accept proposal x^* with probability α
7. update all α_i

Adaptive Metropolis Witin Gibbs

Steps (2/2)

7. update all α_i
8. if accepted: $x' = x^*$
9. repeat for all dimensions (d times)
10. repeat b times

11. after $b \cdot d$ samples adapt for all dimensions i:
12. if $\alpha_i > 0.44 : ls_i = ls_i + \delta(i)$
13. if $\alpha_i < 0.44 : ls_i = ls_i - \delta(i)$
14. $\alpha_i = 0$

15. adaption finished $\Rightarrow$ back to start

Adaptive Metropolis Witin Gibbs

Example . . .

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria
2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results
3. Summary

Outline

1. Introduction
   - Markov Chain Monte Carlo
   - Why adaptive?
   - Criteria

2. Methods
   - Metropolis Hastings
   - Adaptive Metropolis Hastings
   - Adaptive Metropolis Witin Gibbs
   - Regional Metropolis Hastings
   - Results

3. Summary

Model

summer meee