An Introduction to Markov chain Monte Carlo methods

AMATH777 – Stochastic Processes in the Physical Sciences

Tim Dockhorn

University of Waterloo

April 3, 2018

Outline

1. The Monte Carlo method
   - Introduction
   - Example – Approximating π
   - Sampling methods
2. The Markov chain Monte Carlo method
   - Markov chains
   - Example – The random walk Metropolis algorithm

Introduction

Expected value

$$\mathbb{E}[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx$$

Monte Carlo estimate

$$I_N(h) = \frac{1}{N} \sum_{i=1}^{N} h(x^{(i)}) \quad \text{where } \{x^{(i)}\}_{i=1}^{N} \sim f_X(x)$$

with the properties

$$\mathbb{E}[I_N(h)] = \mathbb{E}[h(X)]$$
$$\mathrm{Var}[I_N(h)] = \frac{1}{N} \mathrm{Var}[h(X)]$$

Example – Approximating π

Integrating the area of a unit circle

$$\pi = \int_0^1 \int_0^{2\pi} r dr d\theta = \int_{-1}^1 \int_{-1}^1 \mathbb{1}_A(x,y) dx dy$$
$$= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_A(x,y) f_{X,Y}(x,y) dx dy$$

where

$$\mathbb{1}_A(x,y) := \begin{cases} 1 & \text{if } x^2 + y^2 \leq 1 \\ 0 & \text{otherwise} \end{cases} \qquad f_{X,Y}(x,y) = \begin{cases} \frac{1}{4} & \text{if } |x| \leq 1, |y| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Example – Approximating π

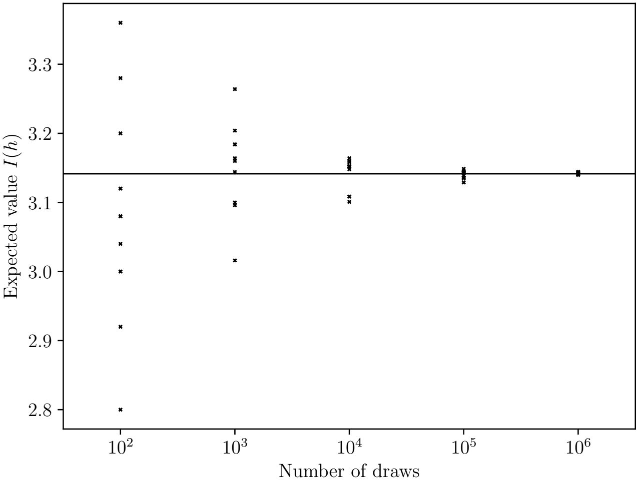
Integrating the area of a unit circle using a Monte Carlo method

$$\pi = 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_A(x, y) f_{X,Y}(x, y) dxdy$$

$$\pi \approx I_N(\mathbb{1}_A(x, y)) = 4 \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{A(x^{(i)}, y^{(i)})}$$

where the tuple $(x^{(i)}, y^{(i)})$ is drawn from the uniform distribution $\mathcal{U}([-1, 1] \times [-1, 1])$.

Example – Approximating π



Sampling methods – How to draw independent and identically distributed samples?

Algorithm 1 Drawing samples from $\mathcal{U}[0,1]$ using a linear congruential generator

1: Choose $x^{(0)} \in \mathbb{R}$ and $a, c, m, N \in \mathbb{N}$
2: $i = 0$
3: **for** $i < N$ **do**
4: $\quad x^{(i+1)} = (ax^{(i)} + c) \mod m$
5: $\{u^{(i)}\}_{i=1}^{N} = \{x^{(i)}/m\}_{i=1}^{N}$

with, e.g., $m = 2^{32}$, $a = 1664525$, $c = 1013904223$, and $0 \leq x^{(0)} < m$ [3].

Sampling methods – How to draw independent and idendtically distributed samples?

Algorithm 2 Inverse transformation method $\left(\text{e.g. } f_X(x) = \lambda e^{-\lambda x}\right)$

1: Draw samples $\{u^{(i)}\}_{i=1}^{N}$ from $\mathcal{U}[0,1]$.
2: $i = 0$
3: **for** $i < N$ **do**
4: $\quad x^{(i)} = F_X^{-1}(u^{(i)})$

How do we draw samples from an arbitrary probability distribution?

Outline

Markov chain

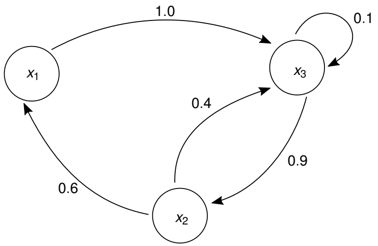


Figure: State transition diagram for a three-state Markov chain [1]

Transition matrix T is defined as $T_{ij} = P(x^{(k)} = x_j | x^{(k-1)} = x_i)$

$$T = \begin{bmatrix} 0.0 & 0.0 & 1.0 \\ 0.6 & 0.0 & 0.4 \\ 0.0 & 0.9 & 0.1 \end{bmatrix}$$

Definition

A Markov chain is a sequence of possible events in which the probability of jumping in one event solely depends on its present state, i.e.,

$$P(x^{(k)} = x_k | x^{(k-1)} = x_{k-1}, x^{(k-2)} = x_{k-2}, ..., x^{(1)} = x_1, x^{(0)} = x_0) = P(x^{(k)} = x_k | x^{(k-1)} = x_{k-1})$$

Markov chain

- If the the state transition graph is both connected (irreducibility) and the chain does not get trapped in cycles (aperiodicity), Markov chain will converge to a stationary distribution $p_N(x)$ with the property that [1]

$$p_{n+1} = p_n T \quad \text{and for "big" } n = N : \quad p_N(x) = p_N(x) T$$

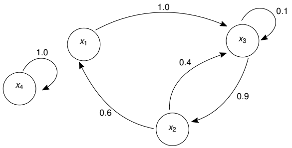


Figure: Markov chain that is not connected

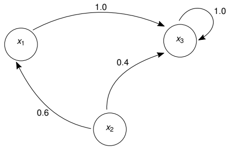


Figure: Markov chain that has a trapping cycle

- Markov chain Monte Carlo sampling methods are Markov chains that have the desired probability distribution $f_X(x)$ as a stationary distribution

Example – The random walk Metropolis algorithm

We try to draw samples from a non-normalized probability distribution

$$f_X(x) = 0.3\exp\big(-0.2x^2\big) + 0.7\exp\big(-0.2(x-10)^2\big)$$

using the proposal distribution $q(x^{(i)}) = \mathcal{N}(x^{(i)}, 100)$.

Algorithm 3 The random walk Metropolis algorithm [2]

```
Choose a x^(0) ∈ supp(f_X(x))
Draw samples {u^(i)}_{i=1}^N from U[0,1]
i = 0
for i < N do
    Draw a sample x* from q(x^(i))
    if u^(i) < min{1, f_X(x*) / f_X(x^(i))} then
        x^(i+1) = x*
    else
        x^(i+1) = x^(i)
```

Example – The random walk Metropolis algorithm

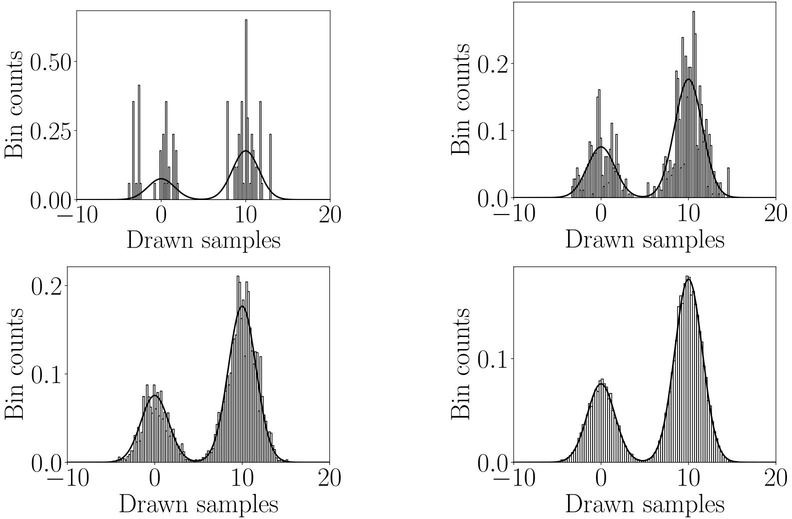


Figure: Markov chain Monte Carlo approximation of the target distribution $f_X(x)$ using $N = \{10^2, 10^3, 10^4, 10^5\}$ draws

[1] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan.
An Introduction to MCMC for Machine Learning.
Machine learning, 50(1-2):5–43, 2003.

[2] N. Metropolis and S. Ulam.
The Monte Carlo Method.
Journal of the American statistical association, 44(247):335–341, 1949.

[3] W. H. Press.
Numerical Recipes 3rd edition: The Art of Scientific Computing.
Cambridge university press, 2007.