The Metropolis-Hastings Algorithm and Its Applications to Computing Integrals in the Orthogonal Group of Matrices.

An Overview

Jacqueline Doan
STATS 9864A
School of Mathematical and Statistical Sciences
Western University
November 11, 2021

Contents

1 Introduction **2**

2 Background **2**
2.1 Orthogonal and Unitary Matrices 2
2.2 Monte-Carlo Markov Chain 3
2.3 Monte-Carlo Markov Chain Algorithm to Generate Random Orthogonal Matrices 4

3 Implementation **4**

4 Analysis **4**

1 Introduction

Random matrix theory is an active branch of modern mathematics, receiving contributions from researchers with diverse backgrounds such as theoretical physics and probability theory. Its applications span from fields that are rich in theory, such as *random graph theory*, to fields with more applications such as *finance*. In this report, we will discuss the Monte Carlo Markov Chain sampling method used to generate random orthogonal and unitary matrices based on the work of W.K. Hastings in 1970. Furthermore, we will use the generated Markov chain to estimate the integral

$$J = \int_{O(m)} f(H)dH$$

where $O(m)$ is the group of orthogonal matrices of size m and dH is an invariant measure.

2 Background

In this section, introductory concepts relevant to this report will be introduced.

2.1 Orthogonal and Unitary Matrices

We will assume the readers are familiar with eigenvalues, eigenvectors, diagonalibility, groups, and elementary probability theory.

A real matrix Q is **orthogonal** if $QQ^T = Q^TQ = I$ where Q^T denotes the transpose of Q and I is the identity matrix. A complex matrix U is **unitary** if $UU^* = U^*U = I$ where U^* is the conjugate transpose of U. An unitary (orthogonal) matrix Q is always diagonalizable. Furthermore, its eigenvectors form an orthonormal basis for $\mathbb{C}^n$ (or an orthogonal basis for $\mathbb{R}^n$ if Q is orthogonal).

The **orthogonal group** (denoted $O(m)$) and the **unitary group** (denoted $U(m)$) under matrix multiplication are defined accordingly, that is

$$O(m) = \{Q \in \mathrm{GL}(n, \mathbb{R}) : Q^TQ = QQ^T = I\} \text{ and}$$
$$U(m) = \{U \in \mathrm{GL}(n, \mathbb{C}) : U^*TU = UU^* = I\}$$

where $GL(m, \mathbb{F})$ is the general linear group, or the group of invertible matrices, of size m over the field $\mathbb{F}$.

Two important classes of random matrices in random matrix theory are the **Guassian Orthogonal Ensemble (GOE)** and the **Guassian Unitary Ensemble (GUE)**. Let us work with the finite dimensional GOE for the rest of this report. A real **Wigner matrix** is a matrix X of size m such that for $i \neq j \in \{1, 2, \ldots, m\}$, $X_{ij} = X_{ji}$ are random variables of some distribution with mean 0 and variance 1, and X_{ii} has some distribution (not necessary the same) with mean 0 and variance 1. If $X \in$ GOE, then $X_{ij} = X_{ji} \sim N(0, 1)$ and $X_{ii} \sim \sqrt{2}N(0, 1)$. Since the GOE contains real and symmetric matrices, the characteristics of the ensemble can be captured by studying the orthogonal group $O(m)$.

We note that $O(m)$ is a Lie group, that is, $O(m)$ is both a group and a differentiable manifold, allowing us to do calculus on it. With this observation, the idea of "taking the integral of a matrix" is now well-defined. Let us first define an **invariant measure**. Suppose (X, Σ) is a measurable space and $f : X \to X$ is a measurable function. A measure μ on (X, Σ) is invariant under f if for all measurable set A in the sigma-algebra Σ,

$$\mu\left(f^{-1}(A)\right) = \mu(A)$$

where $f^{-1}(A)$ is the preimage of A in X. With the unique, translation, and invariant **Haar measure** dH, we can pose the question about calculating the integral

$$J = \int_{O(m)} f(H)dH.$$

In this report, we implement a method in order to numerically evaluate this integral for a given f.

2.2 Monte-Carlo Markov Chain

A **discrete state space Markov chain** is a sequence of random variables $\{X_k\} \in \mathcal{X}$ (finite or countable set) such that

$$P(X_n = x_n | X_1 = x_1, X_2 = x_2, \dots, X_{n-1} = x_{n-1}) = P(_n = x_n | X_{n-1} = x_{n-1})$$

We say a Markov chain is **stationary** if

$$P(X_n = x | X_{n-1} = x')$$

is independent of n. Without loss of generality, let $\mathcal{X} = \mathbb{Z}_{\geq 0}$ and call them *states*. We say π is a **stationary distribution** if $\pi_i \geq 0$, $\sum_i = 1$ and $\pi = \pi P$. Suppose π is a stationary distribution, then

$$\pi = \pi P = (\pi P)P = \pi P^2 = \cdots = \pi P^n$$

The **transition matrix** for a stationary Markov chain is P such that

$$P_{ij} \coloneqq P(X_n = j | X_{n-1} = i)$$

P has non-negative entries and each row sums to 1. Next, we introduce the following notation:

$$\begin{aligned} P_{ij}^{(k)} &\coloneqq P(X_{n+k} = j | X_n = i) \\ f_{ij}^{(k)} &\coloneqq P(X_{n+k} = j, X_{n+k-1} \neq j, \dots, X_{n+1} \neq j | X_n = i) \end{aligned}$$

We have that a Markov chain is **irreducible** if for all i, j, there exists K such that

$$P_{ij}^{(k)} > 0$$

In other words, there is no isolated islands of probability. The i^{th} state of a Markov chain is **transient** if

$$\sum_{k=0}^{\infty} f_{ii}^{(k)} < 1$$

and is **recurrent** if

$$\sum_{k=0}^{\infty} f_{ii}^{(k)} = 1$$

That is, we are guaranteed to get back to any state that we leave. A chain is called **recurrent** if all of its states are **recurrent.** State i of a Markov chain has **period** t if $P_{ii}^{(k)} = 0$ unless

$$k = vt \qquad v \in \mathbb{Z}$$

and t is the largest integer with this property. The state is **aperiodic** otherwise. That is, there is no cyclic structure where we can get back to a state in some set number of transitions (> 1) or a multiple. It follows that a finite-state-space **irreducible** Markov chain is **recurrent**. Define

$$T_{i_0} \coloneqq \min(n : n \geq 1, X_n = i_0)$$

then we have that state i_0 is **recurrent** if

$$P(T_{i_0} < \infty | X_0 = i_0) = 1$$

Furthermore, an **irreducible**, **aperiodic**, and **recurrent** Markov chain, such that the mean recurrence time is finite, that is,

$$\forall i, \qquad E(T_{i_0} | X_0 = i_0) = \sum_{k=0}^{\infty} k f_{ii}^{(k)} < \infty$$

then the chain is **positive recurrent**, or,

$$E(T_{i_0}|X_0 = i_0) = \sum_{k=0}^{\infty} n f_{i_0 i_0}^{(k)} < \infty$$

Furthermore, a stationary distribution for the Markov chain exists, that is, $\exists \pi_j$ with $\pi_j > 0$ and $\sum_j \pi_j = 1$ such that

$$\pi_j = \sum_j \pi_j P_{ij} \qquad \text{Equation of full balance}$$

or in matrix notation, $\pi^T P = \pi^T$ where $\pi = (\pi_1, \pi_2, \dots)^T$.

2.3 Monte-Carlo Markov Chain Algorithm to Generate Random Orthogonal Matrices

Let Q be an orthogonal $n \times n$ matrix. We will generate the sequence of orthogonal matrices $Q(t)$ where $t \in \mathbb{N}^+$.

1. Let $Q(0) = Q_0$ where Q_0 is an orthogonal matrix.
2. At each t, fix $i \neq j$ from $\{1, 2, \dots, n\}$ and $\theta \sim U[0, 2\pi]$
3. Construct the elementary orthogonal matrix $E^{(ij)}(\theta)$ such that

$$E_{\alpha\beta}^{(ij)} = \begin{cases} \cos(\theta) & \text{if } \alpha = \beta = i \\ \sin(\theta) & \text{if } \alpha = i, \beta = j \\ \cos(\theta) & \text{if } \alpha = \beta = j \\ 1 & \text{if } \alpha = \beta \neq i, j \\ 0 & \text{otherwise} \end{cases}$$

4. Choose l from $\{i, j\}$ with probability $1/2$.
5. Construct the matrix $K^{(ij)}(\theta)$ by multiplying the l^{th} row of $E^{(ij)}(\theta)$ by ± 1, each with probability $1/2$.
6. Let $Q(t+1) = K^{(ij)}(\theta) Q(t)$

In order to estimate the integral

$$J = \int_{O(n)} f(Q) dQ$$

for an invariant measure dQ, we calculate the following finite series for large N

$$J \approx \frac{1}{N} \sum_{t=1}^{N} f(Q(t))$$

3 Implementation

4 Analysis