UNIVERSITY OF MANNHEIM
School of Social Sciences
Department of Political Science

Final Paper for Course
Advanced Quantitative Methods in Political Science

On Using the Metropolis-Hastings Algorithm for Data Imputation

Tobias Stenzel
tobias.stenzel@students.uni-mannheim.de

Prof. Thomas Gschwend, Ph.D.

Submission Date: April 19, 2022

Contents

1 Background **1**
1.1 Introduction 1
1.2 Motivation 1
1.3 Markov Chains 1
1.4 The Algorithm 6
1.5 A short proof 6

References **7**

Statutory Declaration

1 Background

1.1 Introduction

The Metropolis-Hastings (MH) algorithm is a method for sampling data points from a probability distribution. It places among the top 10 algorithms with the greatest influence on science and engineering in the 20th century (Beichl and Sullivan 2000). The MH algorithm belongs to the class of Markov chain Monte Carlo (MCMC) methods. In my explanation I assume prior knowledge on Monte Carlo sampling. However, I will explain the basics of Markov Chains. This section is structured as follows. First, I motivate the usage of the MH algorithm. Second, I explain the basics of Markov Chains. Third, I present the algorithm, and finally, I show why it works.

1.2 Motivation

One main application for the MH algorithm is Bayesian inference. Specifically, we want to estimate parameters θ of some probabilistic model f. We have only limited prior knowledge of the distribution of θ and we have a likelihood sample of f. The goal is to estimate the posterior distribution of θ given all information that we have. In practice, we do not have a formal definition of the likelihood but only observations. Therefore, we can only approximate the posterior by numerical integration. This means, we can sample from the posterior and, for example, compute summary statistics like mean and variance of θ. Claassen (2019) assumes a model f parametrized by θ and uses incomplete panel data as the likelihood. Then, he estimates the parameter distribution from the posterior obtained by the MH algorithm. In a final step, he uses the parameter estimates and samples the missing observations from the probabilistic model f.

1.3 Markov Chains

A Markov chain is a stochastic process (over time) with the property that the probability of the realization in the next period depends solely on the realization in the current state and not the complete history. This is called the Markov property. Because discrete Markov chains are much more accessible than their continuous variant, in this chapter we will only look at the discrete case. Formally, the Markov property writes

$$P(X_{t+1}|X_t, X_{t-1}, ..., X_0) = P(X_{t+1}|X_t). \tag{1}$$

Under some conditions, the stochastic process described by a Markov chain converges to a time-invariant probability distribution, i.e. $P(X_{t+k}|X_{t+k-1}) = P(X_t|X_{t-1}), \forall k > 0$. The crucial step for understanding the MH is to see how it samples a Markov Chain that is certain to converge to a stable posterior distribution. Before exploring how the MH algorithm employs these conditions, however, it is necessary to understand them conceptually. To this end, we will use the example depicted by the following graph in Figure 1 that shows the intertemporal transition probabilities between three states representing random events.

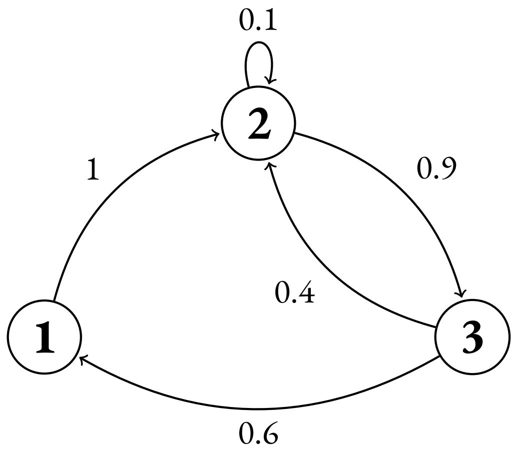

Figure 1: Transition Graph for Markov Chain with 3 states.

This transition graph can be summarized by the $n \times n$ transition matrix T where each element (i, j) represents the probability of moving from state i in period t to state k in period $t + 1$, and where n represents the number of states, i.e $T_{i,j} = P(X_{t+1} = j|X_t = i)$. For our example, we have

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix} \tag{2}$$

1.3.1 Limit Distribution

As touched upon in the previous subsection, interesting questions can be what the probabilities of each state $j \in \{1, ..., s\}$ are after a finite number or infinitely many steps. For this purpose let $\pi(j) = P(X_t = j)$ denote the probability of being in state j in period t. Of course, the probabilities in $t>0 $depend on the probabilities for the the initial state π_0. We can use the law of total probability to calculate the probability of each state for the next period $t = 1$ by

$$P(X_1 = j) = \sum_{i=1}^{3} P(X_1 = j|X_0 = i)\pi_0(i). \tag{3}$$

I.e., to compute the probability of being in state j in $t = 1$, for each initial state i, we multiply its probability $\pi_0(i)$ by the probability of moving from i to state j. This is equivalent to $\pi_1 = \pi_0 T$ in vector notation. Further, we can compute the distributions for distributions locatet in more distant time by repeating the matrix multiplication, e.g, $\pi_2 = \pi_0 TT$, or in general, $\pi_t = \pi_0 T^t$.

Now we are ready to define the limit distribution that describes the probability distribution after infinitely many periods by

$$\pi_\infty = lim_{t\to\infty}\pi_t = lim_{t\to\infty}\pi_0 T^t. \tag{4}$$

We can further ask two additional important questions. First, does a limit distribution exist? And second, is it unique, or in other word, do we have the same limit distribution independent from the realization of the initial state X_0? In our example, there does not only exist a limit distrubtion with $\pi_\infty = (0.2, 0.4, 0.4)$, it is even unique regardless of start distribution π_0!. This means that regardless of the start state, the probability of each state converges to the same number. For the context of the MH algorithm, this is an important property because we always want to compute the same estimates for our parameters θ, regardless of the starting values of our simulation. In the next section, we introduce and simplify conditions that guarantee a unique limit distribution.

1.3.2 Irreducibility, Periodicity and Stationarity

Definition 1.1. A Markov chain is called irreducible if each state is reachable from any other state in a finite number of steps.

Figure 2 shows two times the graph in Figure 1 combined to one Markov chain represented by a bipartite graph. Obviously, this chain is not irreducible because the initial state impacts all future distributions. More precisely, starting in one subgraphs sets the probability of reaching states in the other subgraph to zero. We see that a Markov Chain is only irreducible if there is an indirect link between every pair of states. We also observe that there can be no limit distribution if the Markov Chain is not irreducible.

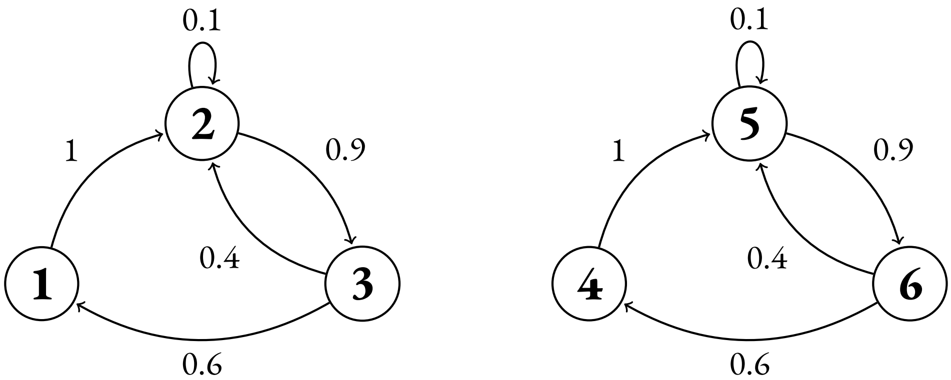

Figure 2: Transition Graph for Irreducible Markov Chain.

Definition 1.2. A state i has a period k if any return to state i must occur in k time periods. A Markov chain is aperiodic if the period of all its states is 1.

Consider the four-state Markov chain in Figure 3 as an illustration for the above definition and suppose we start in state 1. Observe that, independent of the random draw for next period, we will arive again in state 1 after 3 steps. Therefore, state 1 has a period of 3 (as has state 2). Also note that a Markov chain has converged if it is aperiodic (and not vice versa).

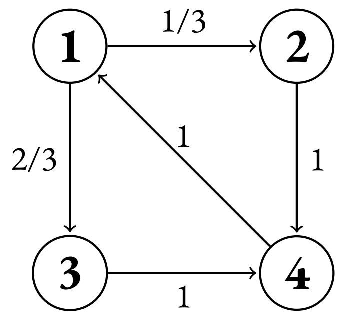

Figure 3: Markov Chain with 3-periodic State A

Definition 1.3. π^* is the stationary distribution of a Markov Chain with Transition matrix T if $\pi^* = \pi^* T$ and π^* is a probability vector.

Verbally, this means that the probability distribution in the next step does not change. Thus, stationarity is a sufficient condition for the statement hat pi^* in some period t has converged. It is, however, a weaker statement than aperiodicity, as it does not require that the "probability flux" out of state i goes into step i in the next period but only that the probability of state i does not change over time independent from which states the "probability flux" into i arrives.

These three definitions are enough to understand the next fundamental theorem.

1.3.3 Basic Limit Theorem

The next theorem defines formally the condition when a Markov Chain converges to a unique distribution.

Theorem 1.1. *If a Markov chain is irreducible and aperiodic with stationary distribution* π^*

Then $\lim_{t\to\infty} P(X_t = i) = \pi_i^*, \forall i$

Therefore, if we want to construct a stable distribution $p(\theta)$ via Markov chains, we need to ensure that it is irreducible and aperiodic with stationary distribution $\pi^* = p(\theta)$. In the next subseciton, we substitute the stationarity condition by a stronger one before we finally come to the MH algorithm.

1.3.4 Reversibility

Definition 1.4. A Markov chain is reversible if there is a probability distribution π over its states such that $\pi(i)T_{ij} = \pi(j)T_{j,i}, \forall i, j$ (reversibility condition).

Theorem 1.2. *A sufficient condition for distribution π^* to be a stationary distribution of a Markov chain with transition matrix T is that it fullfills the reversibility condition.*

Proof. $\sum_i \pi(i)T_{i,j} = \sum_i \pi(j)T_{j,i} = \pi(j)\sum_i T_{j,i} = \pi(j) \implies \pi T = \pi$ □

Reversibility is a stronger condition than stationarity because it requires that the probability flux from i to j is equal to the one from j to i for each possible pair of states. Recall, that stationarity only requires that the probibility flux to one state is equal on aggregate and not that it is symmetric between each pair of states over time.

1.4 The Algorithm

Since we want to estimate model parameters we switch the notation of our markov chain back from x to θ.

Algorithm 1 Metropolis-Hastings algorithm

```
Initialize θ_0
for t ← 0 to T − 1 do
    Sample u ∼ U_[0,1]
    Sample candidate θ* ∼ q(θ*|θ_{t−1})
    if u < min{1, p(θ*)q(θ_t|θ*) / p(θ_t)q(θ*|θ_t)} then
        θ_{t+1} ← θ*
    else
        θ_{t+1} ← θ_t
    end if
end for
```

1.5 A short proof

References

Beichl, Isabel, and Francis Sullivan. 2000. "The Metropolis Algorithm." *Computing in Science & Engineering* 2(1): 65–69.

Claassen, Christopher. 2019. "Estimating Smooth Country–Year Panels of Public Opinion." *Political Analysis* 27(1): 1–20.

Statutory Declaration

Hiermit versichere ich, dass diese Arbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurden. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Veröffentlichungen in gedruckter oder elektronischer Form sind gekennzeichnet. Sämtliche Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen werden kann, wenn die Erklärung nicht erteilt wird.

Mannheim, den __________ ______________________________

Name und Unterschrift

I hereby declare that the paper presented is my own work and that I have not called upon the help of a third party. In addition, I affirm that neither I nor anybody else has submitted this paper or parts of it to obtain credits elsewhere before. I have clearly marked and acknowledged all quotations or references that have been taken from the works of other. All secondary literature and other sources are marked and listed in the bibliography. The same applies to all charts, diagrams and illustrations as well as to all Internet sources. Moreover, I consent to my paper being electronically stores and sent anonymously in order to be checked for plagiarism. I am aware that the paper cannot be evaluated and may be graded “failed” (“nicht ausreichend”) if the declaration is not made.

Mannheim, __________ ______________________________

Name and Signature