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 18, 2022

Contents

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 (insert reference). The MH algorithm belongs to the class of Markov chain Monte Carlo (MCMC) methods. In this section, 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, insert reference) 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, we will look only on the discrete case in this chapter. Formally, the Markov property states that

$$P(X_{t+1}|X_t, X_{t-1}, ..., X_0) = P(X_{t+1}|X_t). \qquad \text{(eq:markov-property)}$$

Under specific condition, the stochastic process described by a Markov chain converges to a time-invariant probability distribution, i.e. $P(X_{t+k}|X_{t+l}) = P(X_t|X_{t-1}) \forall k, l \geq 0, l < k$. The crucial step for understanding the MH is to see how it is able to generate a stable posterior distribution that has converged. Before exploring how the MH algorithm employs these conditions it is thus necessary to understand them conceptually. To this end, we will use the example depicted by the following graph showing the intertemporal transition probabilities between three states:

This transition graph can be summarized by a 3x3 transition matrix T where each element (i, j) represents the probability of moving from state i in period t to state k in period $t + 1$, i.e $T_{i,j} = P(X_{t+1} = j|X_t = i)$. For our example, we have

$$T = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0.1 & 0.9 \\ 0.6 & 0.4 & 0 \end{pmatrix} \qquad \text{(eq:markov-property)}$$

1.3.1 Limit Distribution

As touched in the previous section, one interesting question can be what the probability of each state $j \in \{1, ..., s\}$ is after infinitely many steps. For this purpose let $\pi(j) = P(X_t = j)$ be the probability of being in state j in period t. Of course, the outcome will depend on the initial state π_0 - or the probabilities of the respective state in $t = 0$. Then we can use the law of total probability to calculate the probability of each state for the next period $t = 1$ by

$$P(X_1 = j) = \sum_{i=1}^{3} P(X_1 = j | X_0 = i)\pi_0(i). \qquad \text{(eq:tot-prob)}$$

I.e., to compute the probability of being in state j in $t = 1$, for each initial state i, we multiply its probability 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 more distant distributions by repeating the matrix multiplication, e.g, $\pi_2 = \pi_0 TT$, or more general, $\pi_t = \pi_0 T^t$.

We can now define a limit distribution that describes the probability distribution after infinitely many periods as

$$\pi_\infty = lim_{t\to\infty}\pi_t = lim_{t\to\infty}\pi_0 T^t. \qquad \text{(eq:lim-dist)}$$

Two important question about a well-defined Markov chain is, (i) does a limit distribution exist?, and (ii), is it unique? 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, regardless of the starting values of our simulation. The next sections 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.

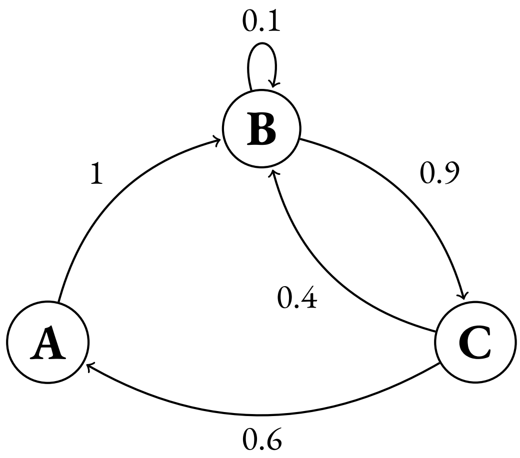

Figure 1: Transition Graph for Markov Chain with 3 states.

Figure 2 shows two times our example Markov chain. Obviously, this Markov chain is not irreducible because the initial state impacts all future distributions because starting in one subgraphs sets the probability of reaching states in the other subgraph to zero. Therefore, a Markov Chain is not irreducible if there is no indicrect link between every pair of states.

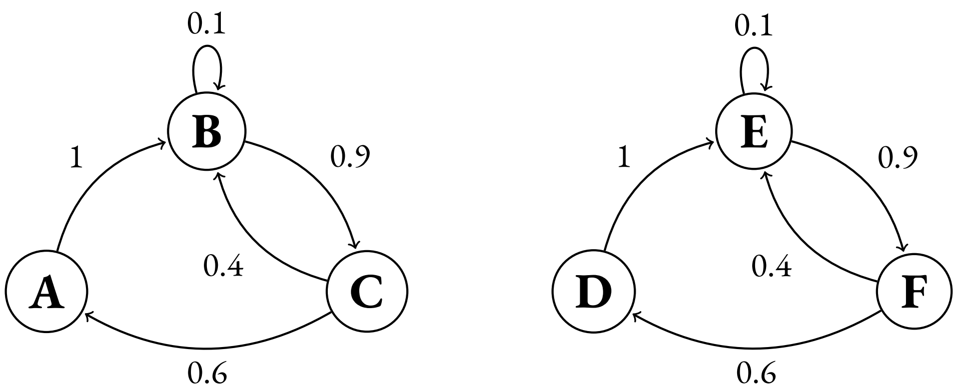

Figure 2: Transition Graph for Irreducible Markov Chain.

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.

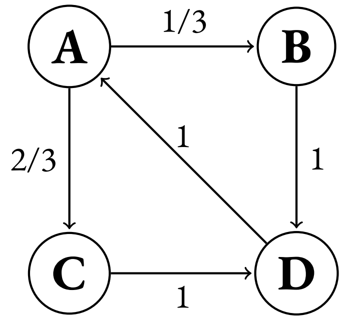

Figure 3: Markov Chain with 3-periodic State A

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

1.3.3 Basic Limit Theorem

1.3.4 Reversibility

1.4 The Algorithm

1.5 A short proof

References

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