THE UNIVERSITY OF CHICAGO

Markov Chain Monte Carlo Method in the 1990s: A case study of the spread of mathematical knowledge

By

Tianli Li

August 2022

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences

Faculty Advisor: Professor Andrew Abbott
Preceptor: Can Merk Kokerer and Claire Watson

1. Introduction

Innovation, diffusion, and application of mathematical knowledge followed entirely different patterns at certain points in history because of the diverging historical context (Struik 1942). In the 20th century, an unprecedented systematization and professionalization in academia took place. Institutions, journals, committees, and academic professionals made up the new academic ecology. In such a highly organized structure, old models of "mathematical advance" may be inadequate.

Numerous studies on scientific and mathematical innovations have been conducted. Social constructivists emphasize the social foundation of advancement in science and technology (Latour 1987; Zilsel 2000). Some studies examine the connection between organizational structures and innovations, attempting to determine the ideal structure to promote innovations and knowledge dissemination in both academia and businesses (Damanpour 1991; Nonaka 1994; Greenhalgh et al. 2004; Shinn 2002). For mathematics, efforts are made to determine the economical, ideological, and political causes of the emergence, evolution, and disagreement of particular mathematical concepts throughout history (Struik 1942; Bloor 1976; MacKenzie 1978; MacKenzie 1981; Hessen 2009). Other works examine diverging cultures and patterns of mathematical innovation in various contexts as they relate to the shift occurring within academia (Lakatos 1976; MacKenzie 1999; MacKenzie 2005; Dick 2011; Dick 2015). Research also focuses on computational algorithms, examining their effects on society (MacKenzie and Spear 2014; MacKenzie 2019). Plenty of social studies have been conducted on a variety of mathematical topics, but few sociological theories focus on explaining the sudden spread of mathematical innovation.

The question of how and why particular mathematical knowledge may see a quick proliferation in contemporary academia remains unanswered. Academic institutions are unprecedentedly well-established in the present day; the academic system, comprised of networks of researchers, academic organizations, journals, and the higher education system, appears to provide a more stable environment for the continuous and effective knowledge production than ever before.

To be sure, big revolutions in pure mathematics, such as the introduction of abstract algebra and set theory, are uncommon in modern post-war academia. But the phenomena of mass innovation and diffusion in the field of applied mathematics are very prominent. In addition to the inner paradigm of mathematical study (proof, refutation, and communication within a limited group of scholars, suggested by Lakatos (1976) and MacKenzie (1999)), the development and spread of applied mathematics depends more on external influences, such as the desire for methodological advances in other fields. Consequently, the development of applied mathematical knowledge, such as the MCMC method that will be addressed in this study, is a somewhat complex process, probably more dependent on numerous social and environmental variables than are advances in pure mathematics.

The Markov chain Monte Carlo (MCMC) method provides a useful example within which to examine this complex process. MCMC is a set of sampling techniques that were firstly established in 1953, stayed mostly undiscovered for nearly forty years, and then grew rapidly in the 1990s. Chemical physics gave birth to MCMC, applied mathematics (statistics) facilitated in its development, and it subsequently expanded to other disciplines, such as the social sciences. By the end of the 20th century, MCMC sampling methods were widely employed in academics. The most straightforward explanation for why and how MCMC spread is simple utility: it is a highly powerful sampling tool. However, useful tools are often ignored, and there still remains the question of the detailed forces leading to adoption of MCMC. There is some historical analysis of the origins of MCMC (Hitchcock 2003; Robert and Casella 2011), but there is little research on the rapid growth and diffusion of MCMC in the 1990s.

This essay describes the history of MCMC and examines its development in the 1990s. Major methods include historical text library research and publishing data analysis. Here, JSTOR, Web of Science, Google Scholar are used as databases. The main goal of this project is to exam the elements that influenced the spread of MCMC in the 1990s. These elements are not unique to MCMC; they continue to play vital roles in the communication and diffusion of various forms of mathematical innovation in current academia. What happened to MCMC in the 1990s demonstrates that the development and transfer of mathematical knowledge can be a complex, multi-faceted process in which conferences, computing power, software, and "explaining" publications all played their roles, giving us a glimpse into how and why specific mathematical knowledge experienced a sudden spread in a well-established modern academic system. The essential result of the analysis is that there is no one dominant force in the spread of such a technique. Spread always depends on the contingent interaction of various forces. Detailed discussion of the example allows us both to list those forces and to show the various ways they interact.

2. MCMC: Overview and History

2.1 A brief introduction to the method

As suggested by its name, Markov chain Monte Carlo (MCMC) refers to a group of algorithms that implement Monte Carlo sampling using a designed Markov chain. These include the Metropolis algorithm, simulated annealing, the Metropolis-Hastings method, Gibbs sampling, simulated annealing, and Hamiltonian Monte Carlo, among others. In these techniques, the precise methods for creating Markov chains vary, but the overall concept of using Markov chains to approach distributions remains the same.

Traditional Monte Carlo methods draw samples independently. In contrast, MCMC has a Markovian property: the next draw is dependent on the previous one in the sampling process. Each sample in MCMC algorithms can be thought of as a state of a Markov chain, and the Markov chain is assumed to have a stationary distribution. As a result, drawing samples can be thought of as moving from one state to another in the chain. Mathematically, we know that if the Markov chain is irreducible and aperiodic, it will eventually reach equilibrium and force the simulation samples to follow the distribution we want to sample (Levin 2017). When compared to the conventional Monte Carlo method, MCMC has many advantages, particularly when simulating higher-dimensional distributions and performing Bayesian statistics. This will be further discussed in the paper, and a more comprehensive illustration is in Gilk et al. (1996).

The most prominent and essential MCMC algorithms are the Metropolis-Hastings (M-H) algorithm and Gibbs sampling. Nicholas Metropolis et al. (1953) developed the M-H algorithm, which uses a form of rejection sampling to generate the Markov chain. Later, Hastings updated this technique (1970). Geman and Geman (1984) developed Gibbs sampling methods; this particular MCMC approach samples from entire conditional distributions and is effective in high-dimensional scenarios[1]. During and after the MCMC boom in the 1990s, a number of other MCMC family members were developed.

1.2 The Birth of the Method

Both MCMC and the Monte Carlo algorithm were created to answer particular physics questions, and MCMC can be seen as an extension of the MC method. Numerous connections between these two methods can be found: both of them were created at Los Alamos National Laboratory[2]; physicist Nicholas Metropolis was heavily involved in the development of both of the two methods[3], even suggested the names (Metropolis 1987)[4]; other authors of the 1953 MCMC breakthrough paper, A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller, were coworkers of von Neumann, Fermi, and Ulam (they were the first to develop and implement the computer-based simulation) at the Los Alamos Lab. If we examine the origins of the two methods, we can see that they are the results of sequential efforts of a group of scientists inventing methods to solve problems on the properties of a collection of particles.

Monte Carlo method was published by Metropolis and Ulam in 1949 as a viable approximation method to tackle problems in physics when an analytical approach fails, despite the fact that the idea of MC had been tested in earlier studies on neutron and gamma-ray diffusion (Goldberger 1948; Metropolis and Ulam 1949). Metropolis and Ulam give an example

[1] Appendix A provides a summary of the two major MCMC approaches.

[2] We can view these methods as products of the post-war Manhattan project.

[3] One of the creators of Metropolis algorithm, Marshall Rosenbluth, claimed that Nick Metropolis only provided the access to the computer MANIAC I, without making mathematical contributions to the method itself at a conference marking the 50th anniversary of the 1953 publication (Gubernatis 2005).

[4] Ulam's uncle gambles in Monte Carlo, and Metropolis suggested using this name to refer to the method (Metropolis 1987).

of a physical condition governed by the Boltzmann equations, which is intractable analytically. An alternative (Monte Carlo) way is to take experimental samples and approach the physical condition by summarizing the result of the experiments. This method consists of two steps: (1) generating random values “with their frequency distribution equal to those which govern the change of each parameter,” and (2) computing "the values of those parameters which are deterministic, i.e., obtained algebraically from the others (Metropolis and Ulam 1949).” The result will end up at the desired phase because of the law of large numbers and asymptotic theorems. Notably, Metropolis and Ulam explain that the mathematical theory behind the repeated sampling process consists of "applications of matrices-like in Markoff chains-and completely specified transformations,” and a similar idea then developed to become the Metropolis algorithm later by Metropolis et al.

In the 1953 Chemical Physics paper entitled "Equation of state calculation by rapid computing machines," the Metropolis method was born to address a problem involving the characteristics of a system of particles. The question waiting to be solved was to calculate the "equilibrium value of any quantity of interest F” in an N-particle system. $\bar{F}$ is calculated by a complex formula: $\frac{\int F * exp(-E/kT) d^{2N}p d^{2N}q}{\int exp(-E/kT) d^{2N}p d^{2N}q}$, where E, the potential energy of the system, is computed by $\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} V(d_{ij})$[5]. It is not practical to do such a complicated integral directly. At first, the Monte Carlo method, which involves randomly placing N particles in a square

---

[5] V is the potential between molecules, and d_{ij} is the minimum distance between particles i and j as defined above (Metropolis et al. 1953)

before solving the equation, was taken into consideration by the authors. However, this technique is likely to select a configuration in which the Boltzmann distribution $exp(E/kT)$ is very small due to the large dimensions (2N) in that situation. This shortcoming severely restricted MC's effectiveness. Then, "a modified Monte Carlo scheme" is proposed as an alternative to solve the issue (Metropolis et al. 1953).

Metropolis et al.'s scheme goes like the following: for particle i (1 ≤ i ≤ N) at (x_i, y_i), propose a new position $(x_i + \sigma\xi_{i1}, y_i + \sigma\xi_{i2})$, where both ξ_{i1} and ξ_{i2} follow a uniform distribution U(-1, 1). Then, accept the proposal with probability $min(1,\ exp(-\Delta E/kT))$, where ΔE is the energy difference between the new configuration and the previous one. If rejected, replicate the value of the previous state as the current state. Having the algorithm, the equilibrium value of F can be approximated by the mean of the samples, $\bar{F} = \frac{1}{M}\sum_{j=1}^{M} F_j$. Although Metropolis et al. did not use the term "Markov chain," they realized the Markovian property by thinking of the samples as states and proved the method's ergodicity[6] to show convergence (Metropolis et al. 1953).

Implementation is the next critical step for this method. Metropolis et al. imagined a system of 224 particles in a unit squared area and used the MANIAC I[7] to run the first MCMC method[8]. The burn-in states were the first 16 runs, and the next 48 to 64 cycles are used to collect

[6] The ability to reach any state from any other.

[7] "MANIAC I" is the acronym of **M**athematical **A**nalyzer **N**umerical **I**ntegrator and **A**utomatic **C**omputer Model **I,** developed under the supervision of Nicklas Metropolis at the Los Alamos.

[8] The parameters they included were the potential between molecules V, and the forbidden distance d_0. See Metropolis et al. (1953).

samples. It took about three minutes per cycle for the Los Alamos MANIAC I, and the method was proven to be feasible after a long series of iterations.

As we examine the very first MCMC method in history, two features stand out to us. First, Metropolis et al. made the a priori assumption that the probability of transitioning from state A to state B is equal to the probability of transitioning from B to A. This is because, in the case of 2N particles, only the positions differed; therefore, "the transition probabilities are equal, otherwise they are zero." Second, the scenario proposed by Metropolis et al. has the additional characteristic of being a higher-dimensional case in which each dimension is cycled through (Metropolis et al., 1953). This iteration approach dealing with high-dimensional cases appears like the other significant MCMC method: Gibbs sampling, which came into existence thirty years later.

Even though Metropolis et al.' s work is one of the most frequently cited papers in statistics and chemical physics, it was not widely known (at least not widely cited) in academia in the middle of the twentieth century. Its current citation data are remarkable: we can find that 48065 citations have been reported by the American Institute of Physics (through June 2022); in the Web of Science core collection, the paper has been cited 24,483 times by works in physics, chemistry, computer science, mathematics, biology, and the social sciences. But prior to 1969, there was never a year where the number of citations exceeded ten[9]. The majority of papers cited Metropolis et al.'s work before 1969 discuss the states of molecules in the field of chemical physics, not the mathematical method. For instance, M.N. Rosenbluth and A. W. Rosenbluth used Los Alamos MANIAC I to test the Metropolis algorithm on the equation of state for three-

[9] According to the WOS databases.

dimensional rigid spheres, demonstrating that the algorithm is a useful tool for solving statistical mechanical problems, although it does not appear to be feasible to "obtain detailed results in transitions regions (Rosenbluth and Rosenbluth 1954)." With the support of the same computer, W. W. Wood and F. R. Parker used the sampling technique to determine the state of equation of Lennard-Jones molecules in a more complicated situation (Wood and Parker 1957). These applications mentioned above were carried out by a small group of scientists who collaborated in the Laboratory to develop mathematical tools for physics and had access to the enormous computer. At least in the 1950s, no physicists who mastered the method were enthusiastic about promoting or employing it in other subjects, and the vast majority of academics lacked access to a dependable computing machine capable of performing such simulations. These limitations restricted the application of the Metropolis algorithm to a narrow niche and prevented its rapid adoption in mathematics and statistics.

1.3 Early Development: Before 1990

In the 1960s, a very limited number of mathematicians were aware of the Metropolis algorithm and the promise of combining Markov chain designing and the Monte Carlo method. For instance, in the chapter 9 of the 1964 book *Monte Carlo Methods*, Metropolis algorithm was introduced as “a method of solving problems in equilibrium statistical mechanics (Hammersley and Handscomb 1964).” This caught the attention of statistician W. K. Hastings, who made significant advancements in the algorithm's development in 1970[10]. Another influential book, *The Monte Carlo Method: The Method of Statistical Trials*, was published in 1966, which also

[10] In Hastings’s 1970 paper, he directly mentioned this book in the introduction part.

contains content about making the Monte Carlo method a Markovian process, though it did not specifically mention the Metropolis algorithm (Buslenko and Shreĭder 1966). In 1965, Australian physicist A.A. Barker proposed a slightly different version of Metropolis algorithm to compute the "radial distribution functions for plasmas over a wide range of temperatures and densities" by sampling from the instrumental density. IBM 1620 and IBM 7090 (Barker, 1965), which were smaller and more accessible than MANIAC I, were employed to test the validity of the procedure. However, these devices were still rare(as expensive as more than 80000 dollars in the 1960s) and sluggish to implement the method[11].

W. K. Hastings' work published on *Biometrika* in 1970 is the next significant stride in the development of MCMC. In the second section of the paper "Dependent Samples", Hastings demonstrated the potential for a non-symmetrical transition probability in the Metropolis algorithm. The original Metropolis algorithm assumes that the acceptance rate of a jump from A to B is dependent on the ratio $\frac{\pi(B)}{\pi(A)}$ [12]. Hastings indicates that the proposed transitional kernel may be asymmetrical (the proposed probability of jumping from A to B $q(A|B)$ differs from that of jumping from B to A $q(B|A)$), so a modified acceptance rate in the algorithm becomes $min(1, \frac{\pi(B)}{\pi(A)} * \frac{q(A|B)}{q(B|A)})$[13] (Hastings 1970; Dunson and Johndrow 2020). With this small but

[11] For instance, a machine cycle of IBM 1620 takes 21 microseconds to run. Source: IBM Archives.

[12] $\pi(x)$ is the desired density that we want to sample from.

[13] Hastings gave two elementary examples, and the second is implemented with the help of IBM 7094 model two, an advanced data processing system in the 1960s (1962 - 1969). (Hastings, 1970; IBM Archives)

significant update, we have the Metropolis-Hastings algorithm that is widely in use today. At the time of writing, the WOS database reports more than 7000 citations for Hastings' paper.

Prior to the revolution of the 1990s, MCMC made noticeable strides in the 1980s. Kirkpatrick et al. published the simulated annealing method, an MCMC method that became popular in image processing starting in the 1980s and was based on the Metropolis algorithm (Kirkpatrick et al. 1983). In their 1984 paper, Geman and Geman proposed the Gibbs sampling method, showed how images and statistical mechanics systems are similar, and provided the method's sound mathematical foundation[14] (Geman and Geman, 1984). Tanner and Wong published a stochastic version of the Expectation-maximization (EM) algorithm to calculate the posterior distribution in statistical inference in 1987. This algorithm embraced a similar ideal to the Gibbs sampling method (Gelfand and Smith 1990; Tanner and Wong 1987). The two most fundamental MCMC techniques were developed by the decade's end, paving the way for a boom in the following ten years.

2. The Revolution

The MCMC revolution of the 1990s includes two aspects. The first was the statistical sampling technique's quick development, and the second was the method's wide use across numerous disciplines. The publications from the 1990s show both. By examining the process of MCMC dissemination, we may determine the roles performed by technology, software, conferences, and exhibiting texts in the rapid spread of a given type of mathematical knowledge.

[14] The idea of the Gibbs sampler relies on the Hammersley-Clifford Theorem, published in the 1970s, which proved that a joint distribution can be determined even if we know by only the full conditional distributions. (Besag 1974; Christian and Casella 2011)

I will begin with a description of the spread by reviewing the publication data and uses of MCMC in the 1990s across many fields.

2.1 Publication Data

I'll start by presenting the publication and citation histories of the important papers that contain the keywords "Markov chain Monte Carlo," "Gibbs sampling," "Metropolis-Hastings," and "simulated annealing." I select the journal articles and book chapters from the JSTOR database because they have been subjected to peer review. I've decided to use 1980 to 2010 as my time frame. Few articles discussed these techniques prior to 1980, and after 2010, things became a little bit more complicated as new technologies like neural networks were developed. The 2010s is also far outside the purview of this paper. Just to be clear, the citation counts and the number of papers published are not always reliable indicators of the acceptance or advancement of a particular body of knowledge, but we can see a clear trend in this MCMC case.

[Figure 1 should be placed here]

Figure 1 shows the number of papers over time that contained various MCMC-related keywords. There are a total of 12,356 papers with these keywords published in the JSTOR data base, with 7825 of them containing "Markov chain Monte Carlo," 1897 containing "Metropolis-Hastings," 3143 containing "Gibbs Sampling," and 3400 containing "Simulated Annealing." We can see from the plot that all four of these keywords were discussed more and more during the 1990s. The growth of the simulated annealing technique began earlier, in the middle of the 1980s, when physicists used the technique in their field after being inspired by Metropolis et al. The terms "Gibbs sampling" and "Metropolis-Hastings" began to appear in the data base more frequently after 1990, and the lines increased gradually since then. The name's unification is another finding. After the mid-1990s, the frequency of the use of "Markov chain Monte Carlo" saw the most rapid growth, and the trend persisted after 2000. It might be as a result of academia accepting the collective name "MCMC" and using the name of the precise algorithm, such as "Gibbs sampling," less frequently.

[Figure 2 should be placed here]

Figure 2 shows how MCMC papers have grown across various disciplines. We continue to use book chapters and journal articles from JSTOR from 1980 to 2010 as the data source in this plot. I list five broad categories that could be useful in illustrating the method's overall evolution and dissemination. These fields fall under applied science, biology, mathematics, physical sciences, and social sciences[15]. In general, the 1990s is a decade of expansion for MCMC in all of the five broad academic fields, particularly "mathematics." The number of publications increased from about 40 per year to about 350 per year in the decade, showing an apparent trend in the method's development. In the first decade of the new century, the application of the MCMC method in biology rose substantially due to the proliferation of Bayesian approaches in biological sciences and the substantial funding in this field[16]. To

[15] Here, “social sciences" includes political science, anthropology, behavioral sciences, communications, population studies, psychology, sociology, law, history, education, economics, and business. “Applied science” includes computer science, electronics, engineering, laboratory techniques, research methods, system science, and technology. “Physical sciences” includes astronomy, chemistry, earth sciences, materials science, and physics. Finally, “mathematics” includes applied math, statistics, pure math, and logic.

[16] There are two journals, BIOMETRICS, BIOMETRIKA, that focus on quantitative biometrics methodology. Though these two journals are under the category of “biological science,” many of the papers from the journals are methodological, not the real applications in experiments or real-world data analysis.

summarize, a significant trend of growth can be seen by looking at the publication data of the five disciplines.

2.2 The Advancement of the MCMC Method in the 1990s

The beginning of the revolution was marked by Gelfand and Smith's (1990) paper, which was followed by a number of significant works that helped to advance the MCMC method both theoretically and practically. In their 1990 paper, Gelfand and Smith reviewed and compared three different Monte Carlo-based methods for numerically estimating the marginal probability distributions: stochastic substitution, the Gibbs sampler, and the sampling-importance-resampling algorithm. The paper has received 3961 citations, according to the WOS database. Additionally, it shows how the three techniques could be used to compute Bayesian posterior densities. In the same year, the Gibbs sampler method for computing the Bayesian marginal posterior with a variety of normal data models was discussed by Gelfand et al. Charles Geyer published a paper in 1991 titled "Markov Chain Monte Carlo Maximum Likelihood", which clearly describes how to perform Maximum Likelihood Estimation (MLE) using MCMC and outlines its advantages over maximum pseudo likelihood estimation (Geyer 1991). It seems that this is the first time that the phrase "Markov chain Monte Carlo" appears to be used in a journal article[17]. Then in his 1992 paper "Practical Markov Chain Monte Carlo", Geyer reviewed some contentious MCMC practices and discussed nonparametric methods for estimating the Monte Carlo error in time-series and operations research literature (citations reach 2473 according to

[17] At that time, only the Gibbs sampling and Metropolis-Hastings algorithm were known to belong to this family. Rather to utilizing MCMC as a collective name, previous works tended to expressly reference the two approaches. Geyer also included the names of the two methods in parentheses under "Markov chain Monte Carlo" in his study.

JSTOR) (Geyer 1992). Four comments of Geyer's (1992) paper were published in the same volume (Geyer 1992; Gelman 1992; Madras 1992; Tierney 1992; Polson 1992). In addition to the MCMCMLE, more sophisticated MCMC methods have been proposed, such as reversible jump Markov chain Monte Carlo[18] and adaptive rejection Gibbs sampling (Gilks and Wild 1992; Green 1995). In the meantime, numerous scholars highlighted the convergence of the MCMC and its applications in Bayesian statistics (Rosenthal 1995; Cowles et al. 1996; Smith and Roberts 1993; Tierney 1994; Geyer 1994; Carlin and Chib 1995; Besag et al. 1995). The influential papers I mentioned are ones that have been cited more than 1000 times in the WOS database. Numerous other important works also played a role in the revolution. To sum up, the MCMC method underwent an unprecedented advancement in the 1990s.

2.3 Spread and Applications

I will then go over how the MCMC method was adopted by the social sciences and biology in the 1990s. Borrowed from statistics and physics, MCMC during that decade became a revolutionary quantitative research approach for statistical sampling (or using sampling to perform Bayesian statistics) in the biological and social sciences[19]. An overview of applications in biology will come first, followed by discussions of sociology, psychometrics, econometrics, and political science in the social sciences.

---

[18] Reversible jump MCMC is an influential algorithm. It enables the MCMC algorithms to work without knowing the number of parameters. The citation data of the original Green's paper on reversible jump MCMC has reached 3281.

[19] The Metropolis-Hastings algorithm and Gibbs sampling are created to handle chemical physics problems. Even though many applications have happened in physics, this is hardly the best example of the interdisciplinary spread of the method. In other subjects, like as the humanities, MCMC has limited utility. Thus, social science and biology are suitable topics for discussion.

As it spreads into biology, MCMC adopts a typical path taken by quantitative methods when they first enter a field: they start with purely methodological discussions before moving on to real applications to address practical issues. The earliest MCMC papers in biological science discuss applicable Bayesian techniques in biological research. Biostatistics pioneer D.G. Clayton published his paper on *Biometrics*, introducing Bayesian inferences (made possible by the MCMC algorithm) in the proportional hazards model[20] (Clayton 1991). Statistician Alan Gelfand, who is also one of the authors of the influential 1990s papers introducing MCMC into statistical academia, wrote a paper with Lynn Kuo on Nonparametric Bayesian Bioassay, applying the MCMC in the Bayesian model in biological science (Gelfand and Kuo, 1991). Vounatsou (1995), Pascual (1996), Mau (1999), Kiuchi et al. (1995), and others have also discussed the use of Bayesian inference in biological methodologies. In addition to supporting Bayesian inference, MCMC was also used to estimate the likelihoods in the analysis of complex genetic models and sample configurations of genes sampled from populations of varying sizes (Griffiths 1994; Thompson 1994). Above are some of the representatives of the application of MCMC to biological science during the 1990s, when 279 papers from JSTOR about MCMC in the field were found to be published[21].

MCMC entered social sciences after 1990, mainly being applied in econometrics and psychometrics. The four traditional social sciences, sociology, political science, economics, and

[20] The proportional hazards model is also known as the "Frailty Model," which is a model for family studies of disease incidence.

[21] Both searches contain the terms "Gibbs sampling," "Gibbs sampler," "Markov chain Monte Carlo," and "Metropolis-Hastings." In the JSTOR search, I include the categories "Biological Sciences," "Botany & Plant Sciences," and "Ecology & Evolutionary Biology" in addition to "biology."

psychology[22], will be my main areas of concentration. There were 287 papers in the four disciplines in the 1990s that discussed or applied the MCMC method[23].

[Figure 3 and Figure 4 should be placed here]

The number of papers discussing the MCMC method in sociology, political science, econometrics, and psychology between 1980 and 2010 is shown in Figures 3 and 4. Figure 3 displays a line plot for each key word, and Figure 4 displays more in-depth trends for each social science. The two plots allow us to see that the trend of growth began in the 1990s and accelerated in the 2000s. Figure 3 shows that Gibbs sampling appears to be the MCMC method that was most frequently used in the 1990s in the social sciences, and it also shows the unification of the name “MCMC”. Figure 4 shows the order of MCMC's popularity in these disciplines: econometrics and psychometrics, followed by political science and sociology. The ranking is roughly in line with how quantitative we perceive these topics to be in general.

Similar to the situation in biological science, the MCMC method was primarily used in econometrics research as a tool for Bayesian statistics. Tobit censored regression models, autoregressive time series models, hierarchical versions of Zellner's SUR models, and Markov mixture models, for example, were all developed using a Bayesian approach by methodologist and econometrician Siddhartha Chib (Chib 1992; 1993; 1996; Chib and Greenberg, 1995). All of the above cases require MCMC to carry out the Bayesian statistical methods. McCulloch and Rossi (1994), Jacquier et al. (1994), and Smith et al. (1996) are other early econometrical studies that used MCMC to conduct Bayesian approaches. Other than Bayesian statistics, simulated

[22] The category “economy" mainly refers to econometrics, and “psychology” mainly refers to psychometrics.

[23] From JSTOR database.

annealing was used in econometrics (Goffe et al. 1994) to find the global optimum in econometrical problems. Gibbs sampling was implemented to estimate the multiple integral that econometricians encountered (Hajivassiliou et al. 1996).

The 1990s saw a change in psychometrical methodology due to Bayesian statistics, which was to some extent made possible by the MCMC method. Model selection, publication bias, nonlinear latent variable models, etc., have all been the subject of Bayesian research (Myung, 1997; Cleary, 1997; Arminger, 1998). The development of the item response theory (IRT) models, a large family of mathematical models explaining the relationship between unobservable attributes and their manifestations, was also aided by the advancement of the MCMC, particularly Gibbs sampling, as well as the Bayesian methods. Examples include the work of Baker (1998) and Hoijtink (1997).

Political science has also produced evidence of the MCMC method's widespread use in the 1990s. Smith (1999) used the MCMC simulation to estimate the Strategically Censored Discrete Choice model after Quinn et al. (1999) implemented MCMC to estimate the multinomial probit (MNP) and multinomial legit (MNL) models of voter choice. Gordon (1996) and Bernard (1997) are two notable examples of political science research that used the Bayesian MCMC method.

Gibbs sampling was first used in sociological journals in two papers that were published in *Sociological Methodology* volume 27 in 1997. The Gibbs sampling is described in one article by Andrew Abbott and Emily Barman (1997), who use the technique to examine subsequence regularities in sociology texts from 1895 to 1965. The other, by Leonhard Knorr-Held and Lugwig Fahrmeir (1997), used MCMC simulation dynamic models for adaptable Bayesian nonparametric analysis of discrete-time or grouped duration data (such as the employment duration data). That volume of the journal also contained a number of critiques of Abbott and Barman's work as well as a response. In three additional papers published on *Sociological Method*, MCMC was mentioned as a potential tool for Bayesian inferences or hierarchical models (Hill 1993, Logan 1998, Raudenbush 1999), but not the main point in these pieces.

According to publishing data, fewer applications of MCMC appeared in sociology and political science in the 1990s than in psychology (psychometrics) and economics (econometrics). The bulk of sociology and political science scholars were less conversant with statistical programming than natural scientists, econometricians, and psychometricians; hence, a new statistical sampling approach (and the availability of Bayesian methods) was not widely employed by sociology and political science researchers. As a result, the incorporation of MCMC into sociology and political science is somewhat idiosyncratic: the process relies substantially on a small number of individuals who paid close attention to new quantitative approaches, but the value of MCMC may not be well recognized in the two fields.

So far, we've discussed the MCMC method's introduction and some of its history prior to the year 2000. This sampling method, which was created by physicists at Los Alamos in 1953, developed over about four decades at a relatively slow rate before experiencing a significant wave of development and proliferation in the 1990s. Previous to the 1990s, there was no indication of adequate contact between physicists who mastered MCMC and researchers in other fields. There were insufficient computing resources and a lack of clear MCMC instructions for a greater number of scholars to participate in the usage and development of MCMC. Then, the 1990s witnessed a transformation.

3. The process of the spread

In this section, I will examine the process of MCMC's spread in the 1990s in greater detail. By studying conferences, technical growth, software advancement, instructional publications, and the desire for Bayesian statistics, we will have a clearer understanding of how a sudden proliferation of mathematical technique in a well-established modern academia ecosystem may arise.

3.1 Conferences and Workshops

Conferences and workshops are excellent venues for academics to exchange ideas and produce new knowledge. Some may contribute significantly to the growth of MCMC. Two applied mathematicians who were present during the method's development in the 1990s, Christian Robert and George Casella, reviewed the history of MCMC from their perspective and released a paper in 2011. Robert and Casella claim that Adrian Smith presented the Gibbs sampler's general characteristics for the first time in June 1989 at a Bayesian workshop in Sherbrooke, Québec. They penned:

> "We still remember vividly the shock induced on ourselves and on the whole audience by the sheer breadth of the method: This development of Gibbs sampling, MCMC, and the resulting seminal paper of Gelfand and Smith (1990) was an epiphany in the world of Statistics."

Two years later, a significant workshop was organized at Ohio State University from February 15–17, 1991. Adrian Smith from Imperial College, London, Prem Goel from Ohio State University, and Alan Gelfand from the University of Connecticut held the event. Topics include theoretical aspects of iterative sampling, posterior simulation and Markov sampling, adaptive sampling, and generalized linear and nonlinear models; each is followed by a section on applications. Numerous discussions and presentations have direct relevance to MCMC. While Nick Polson and Wing-Hung Wong presented their research on Gibbs sampling convergence, Martin Tanner discussed EM, MCEM, DA, and PMDA. "Exploring Posterior Distributions Using Markov Chains" is the title of Luke Tierney's presentation, which later became a frequently cited paper with the same name (Tierney 1994). Research on MCMC was discussed by academics like Donald P. Rubin, Andrew Gelman, Wally Gilks, Adrian Raftery, Charles Geyer, and Elizabeth Thompson, and their papers, German and Rubin (1992), Geyer (1992), Gilks (1992), and Tierney (1994), among others, became influential advancement of MCMC published after the workshop[24].

At the start of the 1990s, MCMC-focused workshops continue to be held. The Royal Statistical Society organized a meeting in May 1992 with the Gibbs sampler and other Markov chain Monte Carlo methods as the main topics of discussion. Participants include early MCMC mathematicians Adrain Smith, Wally Gilks, Julian Besag, and Peter Green[25]. They presented four papers, and each was followed by in-depth discussions (Robert and Casella 2011). The INSERM[26] workshop then took place in Cambridge in 1993, followed by a workshop on Bayesian statistics, MCMC, and the programming language BUGS[27]. As a result of the

---

[24] Robert and Casella attached the agenda of the Ohio workshop as the Appendix of their 2011 paper on the history of MCMC.

[25] From the Report of the Council for the Session 1991-92.

[26] INSERM is the acronym of “Institut national de la santé et de la recherche médicale”, which is the French National Institute of Health and Medical Research.

[27] BUGS stands for “Bayesian inference Using Gibbs Sampling”. This project is focused on building software to implement MCMC method. More detailed discussion will be given later.

workshops, the book *Markov chain Monte Carlo in Practice* was written, which is the first book established specifically on using MCMC in various fields, particularly biostatistics (Lunn et al. 2009; Gilks et al. 1996).

The MCMC pioneers shared and discussed their research projects pertaining to MCMC sampling techniques at these conferences, as shown by the archives that we have been able to locate. It is challenging to draw the conclusion that these were the prerequisites for what transpired in the statistical academic community in the 1990s, but we do note that a large number of the presentations and discussions later evolved into significant papers that marked the development of MCMC and played crucial roles in the method's quick adoption.

3.2 Technological change: advanced devices

To effectively implement MCMC, a dependable and easily accessible device is required. An advancement of the power of chips, the spread of personal computers, and an array of new statistical programming language sets the stage for the spread of MCMC in the 1990s.

Implementing MCMC requires a substantial amount of computational power. In order to complete one iteration of, for instance, the fundamental Metropolis algorithm, it is necessary to compute the transitional kernel of the states and create samples from multiple distributions, which may involve a number of matrix calculations and at least two pseudo-random number generation steps. The “size”, or the complexity bound of each MCMC method varies greatly. It depends on the number of estimated parameters, the proposed distribution, and the number of simulations wanted (Matamoros 2020). Besides the complexity of generating each sample, convergence time is another crucial aspect, since the method must perform more iterations in order to pass the burn-in phases and obtain samples from the required distribution if the Markov chain converges slowly. Partly because of this, the convergence issue has been a big topic since the 1990s. Giving algorithm analysis for various MCMC techniques is beyond reach, but we can have a glimpse through some instances. Frigessi et al (1997).'s research indicates that approximate convergence takes time $O(nlog(n))$ as n (sample size) approaches infinity if the target field satisfies certain spatial mixing conditions. According to Belloni and Chernozhukov (2009), the convergence time of the Metropolis algorithm using Gaussian random walk under the CLT framework with some specific settings[28] is bound by $O(d^2)$ after the bum-in period; the convergence time is $O(d^3Ind)$ combining the bum-in phases, where d is the dimension of the parameter. These studies suggest that the bound of MCMC convergence depends, and so does the execution time. However, it is clear that though MCMC is not an extremely complex algorithm compared with some techniques developed in the past thirty years, it highly relies on big computational power, especially when the dimensionality is high and the Markov chain convergence insufficiently.

Metropolis et al. (1953) had access to the MANIAC at Los Alamos, so they could test the method's applicability. In Metropolis et al. (1953) it took MANIAC, the most advanced computer in 1953, four to five hours to generate a sample (a point on the PA/NkT curve) for the 224-particle system. When Hastings revised the Metropolis Algorithm in 1970, significant progress had been made. Hastings performed his examples on an IBM 7094 ii, a 1960s-era computer that can complete a basic machine operating cycle in 2 microseconds and is equipped with the programming language FORTRAN (IBM Archives; Hastings, 1970). However, the device

[28] The cases in the field of algorithm analysis are technical. For a proof or a more detailed explanation, see (Belloni and Chernozhukov 2009).

was extremely expensive, costing approximately $3.5 million (Vleck, 2022), and its speed was still limited. In 1984, when Geman and Geman introduced the Gibbs sampling for the first time, a 32-bit superminicomputer VAX780 allowed them to conduct more complex high-dimensional samplings (Geman and Geman 1984). In the 1990s, the proliferation and advancement of computing devices provided the material support for MCMC's development and spread.

In the 1990s, the central processing unit experienced a rapid increase in speed and a reduction in size. We can find evidence in the archives of the commonly used CPU speed measurements: processor clock speeds[29], the number of transistors, Millions of Instructions Per Second (MIPS), and Floating-point Operations Per Second (FLOPS). In that decade, the clock speeds of processors increased by more than a factor of ten. For instance, 1990 saw the introduction of IBM's POWER1 with a clock speed of 20–30 MHz. In 1998, POWER3 was released, and its clock speed was 200 MHz (Bayko 2003). In the 1980s and 1990s, the number of transistors followed the pattern described by the well-known Moore's law[30]. For example, the transistor counts of the 1979-released Intel 8088 and 1980-released Inter 8051 are respectively 29,000 and 50,000. Then, around the year 1990, Intel released three new types of chips: Intel i960CA (32 bit with cache), Intel i860 ((32/64-bit, 128-bit SIMD, cache), and Inter 80486 (32 bit with 4 KB cache), with respective transistor counts of 600,000, 1,000,000, and 1,180,235. Intel's Pentium III, which was released in 2000, has a transistor count of 21 million, and newer chips

---

[29] Clock speed is measured in Hz (hertz), KHz(kilohertz), MHz(megahertz), or GHz(gigahertz), which tells the number of cycles a CPU executes per second (P90, Henderson 2009).

[30] "The speed of computers, as measured by the number of transistors that can be placed on a single chip, will double every year or two" (Brenner 1997).

can have up to 42 million (Pentium 4, released by Intel in 2000)[31]. The trend of MIPS growth is also abrupt. 1991's Intel i486 DX, for instance, featured 11.1 MIPS at 33 MHz. Intel's Pentium III processor had 2,054 MIPS at 600 MHz when it was released (MIPS is highly correlated with the number of transistors on a single chip, so it is reasonable.). A detailed graph illustrating the growth is available in Mollick (2006).

The proliferation of personal computers has also contributed to the rise in computing power. From 1990 to 1997, the percentage of households with computers rose from 15.25 to 34.6%, while the percentage of graduate student households with computers rose from 37.2% to 65.6%[32]. This increase benefited from the release of new generations of personal computers, such as the Apple Macintosh II (1987), IBM PS/2 (1988), Macintosh IIx (1988), Macintosh Classic (1990), etc. Apple released its first laptop, Apple Macintosh Portable, in 1989, and IBM produced ThinkPad 700, the first generation of the ThinkPad family, during this time period. We do not know the proportion of MCMC scholars or users that implemented the methods on these newly released devices in the 1990s. However, the big change is manifested in the advancement of hardwares: acceptable computing power was no longer a scarce resource.

The new (or more advanced versions of) programming language also reflects the expansion of computer power as a whole, which therefore afforded statisticians and social

[31] *Microprocessor Hall of Fame*, Intel Corporation, archived from the original on April 6, 2008, retrieved August 11, 2007

[32] Bureau of Labor Statistics, U.S. Department of Labor, *The Economics Daily*, Computer ownership up sharply in the 1990s at https://www.bls.gov/opub/ted/1999/apr/wk1/art01.htm.

scientists with a wider range of computational tools to execute algorithms such as MCMC in the 1990s. Examples including Fortran 90[33], MATLAB[34], STATA,[35] R[36], and SPSS[37].

The overall improvement in the quality of affordable computers and well-designed programming languages had made computer power accessible to more scholars across disciplines, increasing their opportunities to test and implement new algorithms such as MCMC (and use the BUGS packages designed specifically for MCMC).

3.3 The BUGS project

The BUGS project is one of the most notable advancements in software supporting MCMC in the 1990s. Researchers at the Medical Research Council Biostatistics Unit in Cambridge initiated this project to develop a practical statistical package performing MCMC in Bayesian inference in 1989. At the time, mathematicians (statisticians) such as Geman and Geman (1984) had done some work to defend the efficacy of MCMC; some MCMC techniques, like simulated annealing and Gibbs sampling, had been utilized in physics and graph processing.

---

[33] FORTRAN is an IBM-developed compiled imperative programming language. It was formerly the dominant numerical programming language and is still in use today. FORTRAN stands for “Formula Translating System” (Backus, 1998). Fortran 90 is a new generation of FORTRAN that includes a number of updates, such as free-form input, the ability to operate on arrays, operator overloading, etc.

[34] MATLAB was released as a commercial product in 1984 at the Automatic Control Conference in Las Vegas.

[35] STATA was initially developed by William Gould in 1984 and later by Sean Becketti, which initially ran on DOS and was later made available for new generations of PCs and operating systems (Windows and Linux) in the 1990s.

[36] R was created by Robert Gentleman and Ross Ihaka at the Statistics Department of the University of Auckland in 1991 and made available for free in 1995.

[37] SPSS stands for “Statistical Package for the Social Sciences.” It continued to develop new operating system-compatible versions. During the entire decade of the 1990s, the technological field was permeated with modifications.

In addition, a small number of researchers began to recognize the value of stochastic methods for causal inference in Bayesian networks and Bayesian statistics[38] (Pearl 1982; Spiegelhalter 1986; Pearl 1987). With the advancement of MCMC theories and computers, it is time to develop a user-friendly and reliable statistical programming package for MCMC and Bayesian inference using MCMC.

The full name of the project is "Bayesian inference Using Gibbs Sampling," with the catchy acronym "BUGS." Andrew Thomas, who became one of the authors of the 2012 book *The BUGS Book: A Practical Introduction to Bayesian Analysis*, commenced the project following his appointment to the MRC Biostatistics Unit and led the project. Initially, as the name suggested, Gibbs sampling was the only MCMC method considered by the group, and additional algorithms from the MCMC family were gradually added. Andrew Thomas selected Modula-2 as the programming language to implement the package, and the project made rapid progress: a prototype on a personal computer was demonstrated at the 4th Bayesian meeting in Valencia in April 1991, and a Unix version was introduced at the Practical Bayesian Statistics in Nottingham in 1992. In 1993, the project introduced BUGS Version 0.1 for Unix, and in 1995, Version 0.5, a more refined version (Lunn et al., 2009). The 1993 INSERM workshop on MCMC at Cambridge popularized the BUGS, and in 1996, the project began to be jointly developed by the MRC Cambridge Biostatistics Unit and Imperial College School of Medicine at St Mary's, London, and the sampling capabilities of the software were significantly expanded[39]. Observing the impending dominance of the Microsoft Windows operating system, researchers from the two

---

[38] Especially in medical research, and that explains why the project was launched in the Medical Research Council Biostatistics Unit. A more detailed discussion about Bayesian inference is in the next part.

[39] Information is found at https://www.mrc-bsu.cam.ac.uk/software/bugs/. More detailed examples and instructions, see (Spiegelhalter et al, 1996a).

institutes decided to develop a stand-alone Windows version of the BUGS in order to improve the software's performance and reach a larger audience. They used Component Pascal, an object-oriented programming language, to maintain the open-ended nature of the software (Lunn et al., 2009). The result was published in the late 1990s (Lunn et al., 2000) under the name WinBUGS. Remarkably, WinBUGS was able to implement a general-purpose Metropolis–Hastings sampler for updating continuous scalar quantities when the full conditional distribution is neither available in closed form nor log-concave (Lunn et al. 2009). After the year 2000, BUGS continues to evolve, for example, the open-source BUGS software, namely OpenBUGS, the Linux-based version, namely LinBUGS[40], and an interface for spatial modeling, GeoBUGS[41].

According to Lunn et al. (2009), there are more than 30,000 registered users of WinBUGS and more than 100,000 Google hits for the term "WinBUGS" in 2009. Its popularity results from several advantages. First, the software's Gibbs sampling scheme is designed to be applied to any directed acyclic graph[42]. Then, BUGS's advantageous feature is that it makes no distinction between models that will be fitted well and those that will be fitted improperly. In addition, BUGS has the capability to automatically compute the Deviance Information Criterion (DIC), which is akin to an adaptation of the Akaike Information Criterion to Bayesian models that incorporate prior information (Lunn et al. 2009). The success of BUGS led to *The BUGS*

---

[40] University of Cambridge, School of Clinical Medicine, MRC Biostatistics Unit, Software, The BUGS Project, OpenBUGS at https://www.mrc-bsu.cam.ac.uk/software/bugs/openbugs/

[41] University of Cambridge, School of Clinical Medicine, MRC Biostatistics Unit, Software, The BUGS Project, GeoBUGS at https://www.mrc-bsu.cam.ac.uk/software/bugs/thebugs-project-geobugs/

[42] Directed acyclic graph refers to a family of graphs that is directed and with no directed cycles. In Bayesian inference, a Bayesian network is a system of probabilistic events as vertices in a directed acyclic graph. The likelihood of an event can be calculated from the likelihoods of its predecessors in the Graph. BUGS exploits graphical modeling, and its applicability on all kinds of directed acyclic graph enhanced its flexibility greatly. More detailed discussion is in (Spiegelhalter 1998).

Book: A Practical Introduction to Bayesian Analysis, written by Lunn et al. and published in 2012, which provides detailed instructions for using BUGS.

In sum, the project that aimed to design a powerful software for implementing MCMC in Bayesian inference was a success, which increased awareness of Bayesian modeling and contributed to the spread of the MCMC method within academia.

3.4 The "explaining" papers

We frequently credit the papers and authors who first proposed a quantitative method or made creative modifications for its success. They are absolutely necessary for its growth. The MCMC case, however, demonstrates the potency of some papers that introduce, clarify, or contrast the method. Although these works themselves did not substantially change the method, they acted as a bridge between various fields of study. Works by Gelfand and Smith (1990), George and George (1992), and Chib and Greenberg (1995) will be covered here. These are some of the typical ones in the 1990s that were read and cited frequently.

One of the most significant papers at the start of the 1990s MCMC revolution is "Sampling-based Approach to Calculating Marginal Densities" by Gelfand and Smith. Since its publication in 1990, a very large amount of academic works have cited this piece: 22 in 1991 and 728 in the 1990s; up to this point, it has received 3787 citations overall, of which 1859 are from statistics, 569 from biology[43], 320 from economics, and 304 from social sciences mathematical methods[44].

---

[43] There are two categories: "Mathematical Computational Biology" (417 papers) and Biology (152 papers).

[44] According to the WOS database.

Gelfand and Smith examined and contrasted three sampling algorithms that can calculate marginal densities in this paper: the data-augmentation algorithm created by Tanner and Wong (1987), the Gibbs sampler proposed by Geman and Geman (1984), and a type of importance-sampling algorithm created by Rubin (1987). Gelfand and Smith demonstrated the connection between the Gibbs sampler and the data-augmentation algorithm (Substitution Sampling): Data-augmentation algorithm requires the availability of k(k - 1) conditional distributions, which includes all of the full conditionals; Gibbs sampler requires the set of k (the dimension of the variable) full conditional distributions; otherwise, we cannot determine the joint distribution. Gelfand and Smith introduced the three methods and then illustrated their convergence through a series of numerical examples of the sampling techniques' applications in multinomial models, hierarchical models under conjugacy, multivariate normal models, variance component models, normal means models, and an errors-in-variables model.

Later writings frequently cited Gelfand and Smith's paper as the source of their understanding of Gibbs sampling. Consequently, we are confident that the paper was successful in providing a comparative analysis of the three methods, introducing alternatives for intractable marginal densities, and demonstrating Gibbs sampling as a viable option among the three sampling techniques.

With titles like "Explaining the Gibbs Sampler" and "Understanding the Metropolis-Hastings Algorithm," respectively, George and George (1992) and Chib and Greenberg (1995) are more prominent examples of the "explaining" feature. The papers do not aim at theoretical development or methodological innovation. Instead, they concentrate on using descriptive language and in-depth examples to demonstrate the effectiveness of the two MCMC algorithm based on established theories and practices.

George and George (1992) described the steps to perform Gibbs sampling and who the method works. The examples they used include a two-dimensional distribution joint by a Binomial distribution and a Beta distribution, a two-dimensional exponential distribution, and a three-dimensional distribution joint by a Beta, a Binomial, and a Poisson distribution. Each follows a graph showing that Gibbs sampling gives an extremely close set of samples to the ideal target distribution in the above three cases. They illustrated (no a proof) the idea of the Markov chain convergence by a simple two by two matrix case, and briefly summarized, or referred to, some methods to detect and accelerate the convergence.

Similarly, Chib and Greenberg (1995) offered a thorough and very understandable illustration of the Metropolis-Hastings algorithm's theory and provided a step-by-step tutorial for using the method for practical sampling. More importantly, they provided a concise summary of four families of candidate-generating densities for the next state[45] from Metropolis et al. (1953), Hastings (1970), Chib and Greenberg (1994), and Tierney (1994), providing readers with a clear and instructional view of this particularly perplexing and challenging situation of the M-H algorithm.

Although the two papers contributed little to the expansion of the MCMC method's body of knowledge, their organized narratives and understandable illustrations had a significant impact on its dissemination. According to the WOS database, both papers have received numerous citations: Chib and Greenberg (1995) have received 2210, while George and George (1992) have

[45] The probability of generating the next sample given the current sample.

received 1394. Therefore, we cannot disregard the importance of explaining papers in the dissemination of MCMC knowledge.

3.5 Books

In the 1990s, Markov chain Monte Carlo began to appear in more books; some of them are specifically about the MCMC method, while others introduce MCMC as a tool when discussing Bayesian statistics and other related topics. Edited by W.R. Gilks and D.J. Spiegelhalter from the Medical Research Council Biostatistics Unit Cambridge, and S. Richardson from the French National Institute of Health and Medical Research, *Markov chain Monte Carlo in Practice*, published in 1996, is the first and arguably most significant book on MCMC. Mentioned previously, the book originated from the 1993 INSERM workshop on Bayesian statistics using BUGS held in Cambridge, and the three editors are also key figures in the BUGS project. In addition to the three editors, ten eminent scholars contributed to this book, introducing a range of MCMC-related topics, including the concepts of Markov chain, convergence issues, Bayesian statistics, model determination, software implementation, hypothesis testing, hierarchical models, and applications in medical research such as genetics, medical monitoring, and Hepatitis B immunization, etc. A clear and detailed illustration of the method and its applications are written in the book, which made is a widely known guide to MCMC for researchers from various disciplines. The popularity brought citations. At the time of writing, Google scholar reports 12173 citations of *Markov chain Monte Carlo in Practice*, and

759 citations are prior to 2001[46]. In the following year, another treatise on MCMC was published: *Markov chain Monte Carlo: stochastic simulation for Bayesian inference* (Gamerman D.1996). Besides these treatises particularly on MCMC, this method also appeared in earlier books. Tanner's (1993) book mentioned MCMC (Gibbs sampler and Metropolis-Hastings algorithm) as methods to calculate posterior distributions and likelihood functions, whereas Gerhard's (1994) discussed MCMC in image analysis and random fields.

In the late 1990s, MCMC also began to appear in econometric book such as Clemen et al. (1996) and Chi-Lun Cheng (1998). They are evidence of the proliferation of the MCMC methods; further, they served as media to make this algorithm known to scholars of various backgrounds.

4. Bayesian statistics: the demand of a method

As previously stated, the trajectory of MCMC's spread is closely related to the demand for Bayesian statistics. MCMC makes Bayesian inference feasible in many cases, resulting in a paradigm shift in statistical academia and other statistics-driven fields. The significant need for Bayesian inference also acts as a driving force in the development and spread of the MCMC method. This section will provide a brief introduction to Bayesian statistics, and the demand for MCMC as a technique to generate new knowledge in various fields. Little attempt is made to discuss the theory of Bayesian statistics and its applications generally. This extensive topic is covered in greater depth in textbooks such as Lee (2012) and Koch (2007).

---

[46] Google scholar “Markov chain Monte Carlo in Practice”, at https://scholar.google.com/scholar?hl=en&as_sdt=400005&sciodt=0%2C14&cites=14545331270811264086&scipsc=&as_ylo=1000&as_yhi=2000

Bayesian inference is a set of statistical approaches based on a specific interpretation of probability and the use of Bayes' theorem as the main tool for inference. Bayesian statisticians and philosophers argue that human beliefs come with degrees and that probability should be regarded as credence. Following the tenets, Bayesian statisticians treat the credence distribution of the parameter as the prior, and the observed data as conditions. By applying Bayes' theorem, the credence function can be updated to provide the posterior distribution[47]. Bayesian inference has been viewed as an alternative to the frequentist paradigm, which was an easier, more popular, by a flawed method [48].

Before the spread of MCMC in the 1990s, Bayesian inference had a long history of being desired by mathematicians, statisticians, and scientists. The long history of Bayesian statistical inference dates back to Bayes and Price (1764) and Laplace (1814). Then in the twentieth century, scholars like Keynes (1921) and De Finetti (1974) made progress in Bayesian statistical theories, and there were plenty of advancements in the Bayesian approach to psychological research (Edwards et al., 1963), non-parametric problems (Ferguson, 1973), causal inference (Rubin, 1978), etc. However, the frequentist approach had long been the mainstay of statistical reasoning. The establishment of frequentist inference waited until R. A. Fisher, Jerzy Neyman, and Egon Pearson presented their theory of statistical testing in the 1920s and 1930s (Fisher 1925; Neyman and Pearson 1933). Later, this paradigm expanded to become the most widely used quantitative methodology right after its invention and continues to play significant roles in the majority of statistical science-related fields (Lehmann 1993). The frequentist technique is

---

[47] For a more detailed illustration, see Appendix B

[48] For a more detailed illustration, see Appendix C

advantageous because of its simplicity, yet it has serious limitations. Logically, the fact that almost a contradiction exists under the null hypothesis does not imply that the null hypothesis is nearly false (Falk and Greenbaum 1995). Practically, the alpha (rejection criterion) is selected arbitrarily, which leads to the Replication Fallacy. There are also problems with model selection, the inverse probability problem, etc (Gill 1999). As an alternative, Bayesian inference can provide a solution at the level of theory: the probabilistic nature of the parameters in Bayesian inference gives a more comprehensive understanding of the credence distribution, and the use of prior information makes Bayesian inference ideal when the findings of past research are known.

Long after it was desired, Bayesian inference remained unrealistic in many situations without the help of sampling approaches like MCMC. Since the parameters were treated as density functions, the process of acquiring posterior distribution relies on complicated integrations, and sometimes they are difficult to obtain analytically. For high-dimensional posteriors, it is costly or even impossible to perform multiple integrations. An alternative way is to approximate the distribution by sampling, not integrating directly, and MCMC is a great choice[49].

Beginning in the 1990s, the academia acknowledged the potential of MCMC in Bayesian statistics; Gelfand and Smith are frequently credited with its introduction (1990). Prior to Gelfand and Smith, Peal proposed a stochastic simulation-based method for performing causal inference in Bayesian networks (Peal 1987). Gelfand et al. (1990) present a more concentrated discussion on Bayesian statistics, demonstrating Bayesian inference using Gibbs sampling for

[49] In addition to the MCMC approximation, the Laplace method and adaptive quadrature were developed (Smith and Robert 1993; Kass and Raftery 1995). Due to its simplicity and performance, MCMC remained the most widely used Bayesian-supporting algorithm throughout the 1990s.

normal data models, in the same year as Gelfand and Smith's influential introduction. Smith and Robert (1993) reviewed the MCMC (Gibbs Sampling and Metropolis algorithm) in Bayesian inference and provided commentary. Two years later, in 1995, Besag et al. published a review and discussion of the stochastic method in Bayesian statistics. All of these papers are well-cited, and Bayesian methods and Markov chain Monte Carlo (MCMC) have begun to be utilized in a variety of fields.

In contrast to the analytical approximation, MCMC employs a sample-based method that avoids the complex integral. This innovation is revolutionary, and its eventual dominance can be viewed as a paradigm shift in the way the posterior distribution is approached. By constructing a Markov chain whose stationary distribution is the posterior distribution of interest and allowing the Markov chain to converge, MCMC directly generates samples from the posterior distribution of interest. A set of samples can undergo exploratory data analysis, such as quantile summaries, a description of the shape of the distribution (the expectation, the variance, and the moments), and predictive analysis (Smith and Robert, 1993). How to design the transitional kernels of the Markov chain so that their stationary distribution matches the posterior density is a highly technical question. Besag et al. offer a thorough illustration of the method (1995). Prior to the widespread use of simulation-based methods, Bayesian inference was only applicable in the context of so-called conjugate analyses or through the limited use of numerical integration methods (Lunn et al. 2009). With the help of the MCMC sampling method, Bayesian inference is applicable in a variety of situations, including Bayesian Hierarchical Models, Generalized Linear Models, etc.

Academic publications on Bayesian methods over time demonstrate that its greatest growth occurred in the 1990s, when MCMC was developed and disseminated across numerous fields. Prior to 1990, there were fewer than ten publications per year that addressed Bayesian inference. In 1990, ten articles on Bayesian inference were published. The 1991 total is 38, while the 1994 total is 43. In 1998, the data grew significantly and surpassed 100. The trend continued to the present. 2021 saw the publication of 2,213 papers on Bayesian inference, for example. There are currently 20,930 papers in the WoS core collection that cover Bayesian statistics, and more than one-tenth of these papers mention the application of MCMC methods to the Bayesian method[50].

The spread of MCMC and Bayesian statistics is mutually reinforcing. The rising demand for statistical analysis has greatly benefited the growth of applied statistics, demanding the usage of Bayesian reasoning. The novel MCMC approach was the ideal instrument for implementing Bayesian statistics. With the spread of Bayesian statistics, MCMC has a greater chance of being utilized in a variety of fields in the 1990s. It spread in part due to the demand for novel statistical approaches, and it serves as an epistemological tool for generating new knowledge in a wide range of domains.

5. Concluding Remarks

The remarkable evolution and dissemination of the Markov chain Monte Carlo method are illustrations of the spread of mathematical knowledge. As evidenced by the passages, the spread of MCMC occurred in several stages. First, it was created by physicists who faced

---

[50] According to the Web of Science core collection.

challenging integrations calculating particle energy. The Metropolis algorithm was designed as a method for approximating distributions and implemented on the Los Alamos MANIAC I. About two decades after its birth, it still remained largely in the realm of physics, where it was employed by physicists whose concentration was on difficult problems concerning the natural world, leaving insufficient attention for a substantial advancement of the mathematical approach itself. After Hastings improved the Metropolis algorithm, MCMC (even if it was not called this way at the time) attracted greater attention, but the progress of its dissemination was gradual and limited to physics and graph processing. In the 1990s, conferences, explanatory articles, and instructional books facilitated an influx of the MCMC method. In addition, the development in computing power and the proliferation of personal computers contributed to the explosive growth; new statistical programming languages and the BUGS project also offered the essential software and package for implementing MCMC.

The desire for a more philosophically grounded and practically useful statistical paradigm cannot be disregarded in addition to the aforementioned forces that drive the MCMC's diffusion across disciplines. As a simple and effective tool, MCMC rapidly spread into disciplines as a method for executing Bayesian inference, causing a paradigm shift in statistics. The boom of MCMC in the 1990s was a result of the need for Bayesian statistics and the circumstances that encouraged methodological innovation and spread.

With the present division of labor in academia, statistics and applied mathematics did well in the advancement and diffusion of MCMC when MCMC entered the focus of mathematicians in the early 1990s. However, cross-border penetration remained highly contingent, dependent on key people, papers, academic conferences, and supporting technology methods. Due to these contingencies, in the last decade of the twentieth century, MCMC flourished as a mathematical technique within a well-established academic environment through a very rapid diffusion but remained quite silent before.

It is worthwhile to conduct additional mathematical, philosophical, and sociological research on MCMC. For example, the Bayesian nature of MCMC implementation. In many instances, the probability of transition is subjectively determined based on prior knowledge of the situation. The quality of a constructed Markov chain is frequently evaluated by sampling results. However, without an analytical solution of the desired distribution, it can be difficult to validate the result's validity. Some will therefore assert that MCMC is an art. The MCMC story continues beyond the time scope of this paper. It continues to fund research in numerous fields in the twenty first century. Then as machine learning techniques advanced significantly in the 2010s, MCMC began to be replaced by other algorithms. The tensions and iterations between the various algorithms are the subject of a separate narrative.

In addition, the spread of mathematical innovation remains a thought-provoking problem. The narrative of Markov chain Monte Carlo (MCMC) here provides a glimpse of the quick transmission of specific mathematical knowledge; however, it is unrealistic to list and analyze all key criteria for the success of all mathematical breakthroughs in a single case. As was previously said, the process of the production and transmission of distinct bodies of knowledge may take radically divergent routes. There are still many excellent stories to be unearthed.

Acknowledgment

I would like to acknowledge and give my sincere gratitude to my advisor Professor Andrew Abbott. His guidance has carried me through all stages of this project and brought me into a wider space of social science. He has also provided me an example of an innovative, erudite, and kind-hearted thinker who is passionate about knowledge. The experience working with him will continue to be a priceless treasure in my future intellectual journey.

I would also like to give my thanks to Professor Nicholas Polson, Yanyan Sheng, James Evans, Adrain Johns, Michael Castelle, Donald Mackenzie, and my preceptors Can Merk Kokerer and Claire Watson for their insightful suggestions during my year at the University of Chicago.

Finally, thank my parents. It might not be an easy decision for them to support me to make the transition to social science. I feel very lucky that they can always stand by me and give me trust and love.

Reference:

Abbott, Andrew, and Emily Barman. "Sequence Comparison via Alignment and Gibbs Sampling: A Formal Analysis of the Emergence of the Modern Sociological Article." *Sociological Methodology* 27 (1997): 47–87. http://www.jstor.org/stable/271098.

Arminger, G. A. "Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm." *PSYCHOMETRIKA* 63, no. 3 (1998): p.271-300 ISSN: 0033-3123

Backus, John. "The History of Fortran I, II, and III" (PDF). *IEEE Annals of the History of Computing*. (October–December 1998). **20** (4): 68–78.

Baker, FB. "An investigation of the item parameter recovery characteristics of a Gibbs sampling procedure." *APPLIED PSYCHOLOGICAL MEASUREMENT* 22, no. 2 (1998): p.153-169 ISSN: 0146-6216

Barker, A. A."Monte Carlo Calculations of the Radial Distribution Functions for a Protonelectron Plasma," *Australian Journal of Physics*, 18, (1965) 119-133.

Bayes, T., and R. Price. "An Essay Towards Solving a Problem in the Doctrine of Chances." *Philos*. *Trans*. *R*. *Soc*. *Lond*. (1764). **53** 370–418.

Bayko. John, "Great Microprocessors of the Past and Present (V 13.4.0)". (2003). https://www.cpushack.com/CPU/cpu.html.

Bernard, Jean-Thomas, Stephen Gordon, and Josee Tremblay. "Electricity Prices and Elections in Quebec." *The Canadian Journal of Economics / Revue Canadienne d'Economique* 30, no. 3 (1997): 505–25. https://doi.org/10.2307/136231.

Belloni, Alexandre, and Victor Chernozhukov. "On the Computational Complexity of MCMC-Based Estimators in Large Samples." *The Annals of Statistics* 37, no. 4 (2009): 2011–55. http://www.jstor.org/stable/30243694.

Besag, Julian. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society. Series B (Methodological)* 36, no. 2 (1974): 192–236. http://www.jstor.org/stable/2984812.

Besag, Julian, Peter Green, David Higdon, and Kerrie Mengersen. "Bayesian Computation and Stochastic Systems." *Statistical Science* 10, no. 1 (1995): 3–41. http://www.jstor.org/stable/2246224.

Brenner, A. E. "Moore's Law." *Science.* (1997). *275*(5306), 1551–1551. http://www.jstor.org/stable/2893682

Bureau of Labor Statistics, U.S. Department of Labor, *The Economics Daily*, Computer ownership up sharply in the 1990s at https://www.bls.gov/opub/ted/1999/apr/wk1/art01.htm (visited *June 24, 2022*).

Buslenko, N. P., and I͡U. A. Shreĭder. *The Monte Carlo Method: The Method of Statistical Trials*. Oxford, New York: Pergamon Press, (1966).

Carlin, Bradley P., and Siddhartha Chib. "Bayesian Model Choice via Markov Chain Monte Carlo Methods." *Journal of the Royal Statistical Society.* Series B (Methodological) 57, no. 3 (1995): 473–84. http://www.jstor.org/stable/2346151.

Casella, George, and Edward I. George. "Explaining the Gibbs Sampler." *The American Statistician* 46, no. 3 (1992): 167–74. https://doi.org/10.2307/2685208.

Chan, K., and C. Geyer. Discussion of “Markov chains for exploring posterior distribution.” *Ann. Statist*. (1994). **22** 1747– 1758.

Cheng, Chi-Lun. and John W. Van Ness, et al. *Econometrics in Theory and Practice: Festschrift for Hans Schneeweiss*. Heidelberg ; New York: Physica-Verlag, 1998.

Chib, Siddhartha. “Calculating posterior distributions and modal estimates in Markov mixture models.” *Journal of Econometrics*. Volume 75, Issue 1. (1996). Pages 79-97. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(95)01770-4.

Chib, Siddhartha. “Bayes regression with autoregressive errors: A Gibbs sampling approach.” *Journal of Econometrics*. Volume 58, Issue 3. (1993). Pages 275-294. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(93)90046-8.

Chib, Siddhartha. “Bayes inference in the Tobit censored regression model.” *Journal of Econometrics*. Volume 51, Issues 1–2. (1992). Pages 79-99. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(92)90030-U.

Chib, Siddhartha., and Edward Greenberg. “Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models.” *Journal of Econometrics*. Volume 68, Issue 2. (1995). Pages 339-360. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(94)01653-H.

Chib, Siddhartha., and Edward Greenberg. “Understanding the Metropolis-Hastings Algorithm.” *The American Statistician* 49, no. 4 (1995): 327–35. https://doi.org/10.2307/2684568.

Chib, Siddhartha, and Edward Greenberg. “Markov Chain Monte Carlo Simulation Methods in Econometrics.” *Econometric Theory* 12, no. 3 (1996): 409–31. http://www.jstor.org/stable/3532527.

Clayton, David G. “A Monte Carlo Method for Bayesian Inference in Frailty Models.” *Biometrics* 47, no. 2 (1991): 467–85. https://doi.org/10.2307/2532139.

Cleary, RJ. “An application of Gibbs sampling to estimation in meta-analysis: Accounting for publication bias.” *JOURNAL OF EDUCATIONAL AND BEHAVIORAL STATISTICS* 22, no. 2 (1997): p.141-154 ISSN: 1076-9986.

Clemen, Robert T., Steven K. Jones and Robert L. Winkler --, et al. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner*. New York: Wiley, 1996.

Cowles, Mary Kathryn, and Bradley P. Carlin. “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.” *Journal of the American Statistical Association* 91, no. 434 (1996): 883–904. https://doi.org/10.2307/2291683.

Damanpour, F. “Organizational innovation - a meta-analysis of effects of determinants and moderators.” *ACADEMY OF MANAGEMENT JOURNAL* 34, no. 3 (1991): p.555-590 ISSN: 0001-4273

De Finetti, Bruno. *Theory of Probability: A Critical Introductory Treatment.* London, New York: Wiley, 1974.

Dempster, A. P. “A Generalization of Bayesian Inference.” *Journal of the Royal Statistical Society.* Series B (Methodological) 30, no. 2 (1968): 205–47. http://www.jstor.org/stable/2984504.

Dick, Stephanie. “AfterMath: The Work of Proof in the Age of Human–Machine Collaboration.” *Isis* 102, no. 3 (2011): 494–505. https://doi.org/10.1086/661623.

Dick, Stephanie. “Of Models and Machines: Implementing Bounded Rationality.” *Isis.* 106:3, 623-634 (2015).

Dunson, D. B., and J. E. Johndrow. “The Hastings algorithm at fifty." *Biometrika*, Volume 107, Issue 1, March (2020), Pages 1–23, https://doi.org/10.1093/biomet/asz066.

Edwards, W., H. Lindman, and L. J. Savage. “Bayesian statistical inference for psychological research.” *Psychological Review* 70, no. 3 (1963): p.193-242 ISSN: 0033-295X.

Efron, B. “Why Isn’t Everyone a Bayesian?” *The American Statistician* 40, no. 1 (1986): 1–5. https://doi.org/10.2307/2683105.

Fahrmeir, Ludwig, and Leonhard Knorr-Held. “Dynamic Discrete-Time Duration Models: Estimation via Markov Chain Monte Carlo.” *Sociological Methodology* 27 (1997): 417–52. http://www.jstor.org/stable/271112.

Falk, R., and C. W Greenbaum. "Significance Tests Die Hard." *Theory and Psychology.* (1995). 5: 396-40.

Ferguson, TS. “Bayesian analysis of some nonparametric problems." *ANNALS OF STATISTICS* 1, no. 2 (1973): p.209-230 ISSN: 0090-5364.

Fisher, R. A. *Statistical Methods for Research Workers.* Edinburgh, Oliver & Boyd, 1925.

Fisher, Ronald. “Statistical Methods and Scientific Induction.” *Journal of the Royal Statistical Society*. Series B (Methodological) 17, no. 1 (1955): 69–78. http://www.jstor.org/stable/2983785.

Frigessi, Arnoldo, Fabio Martinelli, and Julian Stander. “Computational Complexity of Markov Chain Monte Carlo Methods for Finite Markov Random Fields.” *Biometrika* 84, no. 1 (1997): 1–18. http://www.jstor.org/stable/2337551.

Gamerman, Dani. *Markov chain Monte Carlo : stochastic simulation for Bayesian inference*. London ; New York : Chapman & Hall, 1997.

Gelfand, Alan E., Susan E. Hills, Amy Racine-Poon, and Adrian F. M. Smith. “Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling.” *Journal of the American Statistical Association* 85, no. 412 (1990): 972–85. https://doi.org/10.2307/2289594.

Gelfand, Alan E., and Adrian F. M. Smith. “Sampling-Based Approaches to Calculating Marginal Densities.” *Journal of the American Statistical Association* 85, no. 410 (1990): 398–409. https://doi.org/10.2307/2289776.

Gelfand, Alan E., and Lynn Kuo. “Nonparametric Bayesian Bioassay Including Ordered Polytomous Response.” *Biometrika* 78, no. 3 (1991): 657–66. https://doi.org/10.2307/2337035.

Gelman, Andrew, and Christian P. Robert. “‘Not Only Defended But Also Applied’: The Perceived Absurdity of Bayesian Inference.” *The American Statistician* 67, no. 1 (2013): 1–5. http://www.jstor.org/stable/24590337.

Gelman, Andrew, and Donald B. Rubin. “[Practical Markov Chain Monte Carlo]: Rejoinder: Replication without Contrition.” *Statistical Science* 7, no. 4 (1992): 503–11. http://www.jstor.org/stable/2246105.

Gelman, A. and Rubin, D. “Inference from iterative simulation using multiple sequences (with discussion).” *Statist. Sci.* (1992). **7** 457–511.

Geman, S., and D. Geman. “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images.” *IEEE Trans Pattern Anal Mach Intell.* (1984). Jun;6(6):721-41. doi: 10.1109/tpami.1984.4767596. PMID: 22499653.

Geyer, Charles J. “Markov Chain Monte Carlo Maximum Likelihood.” *Interface Foundation of North America*. (1991). Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/58440.

Geyer, Charles J. “Practical Markov Chain Monte Carlo.” *Statistical Science* 7, no. 4 (1992): 473–83. http://www.jstor.org/stable/2246094.

Geyer, Charles J., and Elizabeth A. Thompson. “Constrained Monte Carlo Maximum Likelihood for Dependent Data.” *Journal of the Royal Statistical Society. Series B (Methodological)* 54, no. 3 (1992): 657–99. http://www.jstor.org/stable/2345852.

Gill, Jeff. “The Insignificance of Null Hypothesis Significance Testing.” *Political Research Quarterly* 52, no. 3 (1999): 647–74. https://doi.org/10.2307/449153.

Gilks, W. “Derivative-free adaptive rejection sampling for Gibbs sampling.” In *Bayesian Statistics 4* (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.) 641–649. Oxford Univ. Press, Oxford. (1992).

Gilks, W. R., and P. Wild, “Adaptive Rejection Sampling for Gibbs Sampling.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 41, no. 2 (1992): 337–48. https://doi.org/10.2307/2347565.

Gilks, W. R. , S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London. (1996).

Goffe, William L., Gary D. Ferrier, and John Rogers. “Global optimization of statistical functions with simulated annealing.” *Journal of Econometrics.* Volume 60, Issues 1–2. (1994). Pages 65-99. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(94)90038-8.

Goldberger, M. L. “The Interaction of High Energy Neutrons and Heavy Nuclei.” *Phys. Rev*. (1948). **74**, 1269.

Gordon, Stephen. “Using Mixtures of Flexible Functional Forms to Estimate Factor Demand Elasticities.” *The Canadian Journal of Economics / Revue Canadienne d’Economique* 29, no. 3 (1996): 717–36. https://doi.org/10.2307/136259.

Greenhalgh, Trisha., Glenn Robert, Fraser MacFarlane, Paul Bate, and Olivia Kyriakidou. “Diffusion of Innovations in Service Organizations: Systematic Review and Recommendations.” *The Milbank Quarterly* 82, no. 4 (2004): 581–629. http://www.jstor.org/stable/4149085.

Green, Peter J. “Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination.” *Biometrika* 82, no. 4 (1995): 711–32. https://doi.org/10.2307/2337340.

Griffiths, R. C., and Simon Tavare. “Sampling Theory for Neutral Alleles in a Varying Environment.” *Philosophical Transactions: Biological Sciences* 344, no. 1310 (1994): 403–10. http://www.jstor.org/stable/56112.

Gubernatis, J. “Marshall Rosenbluth and the Metropolis algorithm.” *Physics of Plasmas.* (2005). 12. 057303-057303. 10.1063/1.1887186.

Haigh, T., M. Priestley and C. Rope, "Los Alamos Bets on ENIAC: Nuclear Monte Carlo Simulations, 1947-1948," in IEEE Annals of the History of Computing, vol. 36, no. 3, pp. 42-63, July-Sept. (2014). doi: 10.1109/MAHC.2014.40.

Hajivassiliou, Vassilis., Daniel McFadden, and Paul Ruud. "Simulation of multivariate normal rectangle probabilities and their derivatives theoretical and computational results." *Journal of Econometrics.* Volume 72, Issues 1–2. 1996. Pages 85-134. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(94)01716-6.

Hammersley, J.M., and D.C. Handscomb. Problems in Statistical Mechanics. In: Monte Carlo Methods. Monographs on Applied Probability and Statistics. Springer, Dordrecht. (1964). https://doi.org/10.1007/978-94-009-5819-7_9.

Hastings, W. K. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57, no. 1 (1970): 97–109. https://doi.org/10.2307/2334940.

Henderson, H. "Encyclopedia of Computer Science and Technology". (2009). Facts on file science library.

Hessen, B. "The Social and Economic Roots of Newton's *Principia*." In: Freudenthal, G., McLaughlin, P. (eds) The Social and Economic Roots of the Scientific Revolution. Boston Studies in the Philosophy of Science, vol 278. (2009). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-9604-4_2.

Hitchcock, David B. "A History of the Metropolis-Hastings Algorithm." *The American Statistician* 57, no. 4 (2003): 254–57. http://www.jstor.org/stable/30037292.

Hoijtink, H. "A multidimensional item response model: Constrained latent class analysis using the Gibbs sampler and posterior predictive checks." *PSYCHOMETRIKA.* 62, no. 2 (1997): p.171-189 ISSN: 0033-3123.

IBM Archives/ Exhibits/ IBM Mainframes/ Mainframes reference room/ Mainframes product profiles/ 1620 Data Processing System. https://www.ibm.com/ibm/history/exhibits/mainframe/mainframe_PP1620.html

IEEE 1003.9-1992 - IEEE Standard for InformationTechnology - POSIX(R) FORTRAN 77 Language Interfaces - Part 1: Binding for System Application Program Interface (API). IEEE. Retrieved 24 November 2018.

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi. "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business & Economic Statistics* 12, no. 4 (1994): 371–89. https://doi.org/10.2307/1392199.

Kass, Robert E., and Adrian E. Raftery. "Bayes Factors." *Journal of the American Statistical Association* 90, no. 430 (1995): 773–95. https://doi.org/10.2307/2291091.

Keynes, J. M. *A Treatise on Probability*, London: Macmillan. (1921).

Kirkpatrick, S., C. D. Jr. Gelatt, and M. P. Vecchi. "Optimization by simulated annealing." *Science* (1983).220 671–680. MR0702485.

Kiuchi, Amy S., J. A. Hartigan, Theodore R. Holford, Pablo Rubinstein, and Cladd E. Stevens. "Change Points in the Series of T4 Counts Prior to AIDS." *Biometrics* 51, no. 1 (1995): 236–48. https://doi.org/10.2307/2533329.

Koch, Karl-Rudolf. *Introduction to Bayesian Statistics*, Berlin ; New York : Springer, c2007.

Laplace, Pierre-Simon. *A Philosophical Essay on Probabilities*. New York, Wiley, 1902.

Lakatos, Imre. *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge ; New York: Cambridge University Press, 1976.

Latour, Bruno. *Science in Action*. Cambridge, Mass.: Harvard University Press. (1987).

Lee, Peter M. *Bayesian Statistics : an Introduction,* Chichester, West Sussex ; Hoboken, N.J. : Wiley, 2012.

Lehmann, E. L. “The Fisher, Neyman-Pearson Theories of Testing Hypotheses: One Theory or Two?” *Journal of the American Statistical Association* 88, no. 424 (1993): 1242–49. https://doi.org/10.2307/2291263.

Levin, David A., Yuval Peres, and Elizabeth L. Wilmer. Markov Chains and Mixing Times. (2017). http://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf.

Lorenz, Lori L., and R. Michael O’Mara. (1990). Windows 3 Companion. Microsoft Press. **ISBN-13 :** 978-0936767192.

Lunn, D.J., A. Thomas, N. Best, and D. Spiegelhalter. “WinBUGS — a Bayesian modelling framework: concepts, structure, and extensibility.” *Statistics and Computing*, (2000). **10**:325–337.

Lunn, David., David Spiegelhalter, Andrew Thomas, and Nicky Best. “The BUGS project: Evolution, critique and future directions." *Statistics in Medicine.* (2009). 28. 3049-67. 10.1002/sim.3680.

MacKenzie, Donald. “Statistical Theory and Social Interests: A Case-Study.” *Social Studies of Science* 8, no. 1 (1978): 35–83. http://www.jstor.org/stable/284856.

MacKenzie, Donald. *Statistics in Britain, 1865-1930: The Social Construction of Scientific Knowledge.* Edinburgh: Edinburgh University Press, 1981.

MacKenzie, Donald. “How Algorithms Interact: Goffman’s ‘Interaction Order’ in Automated Trading.” *Theory, Culture & Society* 36, no. 2 (March 2019): 39–59. https://doi.org/10.1177/0263276419829541.

MacKenzie, Donald. “Computing and the Cultures of Proving.” *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 363, no. 1835 (2005): 2335–50. http://www.jstor.org/stable/30039731.

MacKenzie, Donald. “Slaying the Kraken: The Sociohistory of a Mathematical Proof.” *Social Studies of Science* 29, no. 1 (1999): 7–60. http://www.jstor.org/stable/285445.

MacKenzie, Donald, and Taylor Spears. “‘The Formula That Killed Wall Street’: The Gaussian Copula and Modelling Practices in Investment Banking.” *Social Studies of Science* 44, no. 3 (2014): 393–417. http://www.jstor.org/stable/43284238.

Madras, Neal. “[Practical Markov Chain Monte Carlo]: Comment.” *Statistical Science* 7, no. 4 (1992): 488–89. http://www.jstor.org/stable/2246097.

Madey, G., X. Xiang, S. Cabaniss, and Y. Huang, "Agent-Based Scientific Simulation" in *Computing in Science & Engineering*, vol. 2, no. 01, pp. 22-29, 2005. https://royalsociety.org/people/adrian-smith-12298/.

Matamoros, Izhar Asael Alonzo. "An introduction to computational complexity in Markov Chain Monte Carlo methods." *arXiv preprint arXiv:2004.07083* (2020).

Mau, Bob, Michael A. Newton, and Bret Larget. “Bayesian Phylogenetic Inference via Markov Chain Monte Carlo Methods.” *Biometrics* 55, no. 1 (1999): 1–12. http://www.jstor.org/stable/2533889.

McCulloch, R., and P. E. Rossi. “An exact likelihood analysis of the multinomial probit model.” *Journal of Econometrics*. Volume 64, Issues 1–2. (1994). Pages 207-240. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(94)90064-7.

Metropolis, Nick. “The beginning of the Monte Carlo method.” *.Los Alamos Science Special Issue.* (1987). https://library.lanl.gov/cgi-bin/getfile?00326866.pdf。

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, and A. H. Teller. “Equation of state calculation by fast computing machines.” J. *Chem. phys.,* **21,** No. 6, 1087-1092 (1953).

Microprocessor Hall of Fame, Intel Corporation, archived from the original on April 6, 2008, retrieved August 11, 2007

Mollick, E. "Establishing Moore's Law," in *IEEE Annals of the History of Computing*, vol. 28, no. 3, pp. 62-75, (July-Sept. 2006), doi: 10.1109/MAHC.2006.45.

Myung, IJ. “Applying Occam's razor in modeling cognition: A Bayesian approach.” *PSYCHONOMIC BULLETIN & REVIEW* 4, no. 1 (1997): p.79-95 ISSN: 1069-9384.

Neyman, J., and E. S. Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses.” Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 231 (1933): 289–337. http://www.jstor.org/stable/91247.

Nonaka, Ikujiro. “A Dynamic Theory of Organizational Knowledge Creation.” *Organization Science* 5, no. 1 (1994): 14–37. http://www.jstor.org/stable/2635068.

Norris, J. R. *Markov Chains*. Cambridge [England] ; New York: Cambridge University Press, 1997.

Pascual, Miguel A., and Peter Kareiva. “Predicting the Outcome of Competition Using Experimental Data: Maximum Likelihood and Bayesian Approaches.” *Ecology* 77, no. 2 (1996): 337–49. https://doi.org/10.2307/2265613.

Pearl J. “Evidential reasoning using stochastic simulation.” *Artificial Intelligence* 1987; **32**:245–257.

Polson, Nicholas G. “[Practical Markov Chain Monte Carlo]: Comment.” *Statistical Science* 7, no. 4 (1992): 490–91. http://www.jstor.org/stable/2246098.

“Preliminary Program of the 1994 Annual Meeting.” *PS: Political Science and Politics* 27, no. 2 (1994): 353–496. http://www.jstor.org/stable/420297.

Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. “Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models.” *American Journal of Political Science* 43, no. 4 (1999): 1231–47. https://doi.org/10.2307/2991825.

“Report of the Council for the Session 1991-92.” Journal of the Royal Statistical Society. Series A (Statistics in Society)155, no. 3 (1992): 421–55. http://www.jstor.org/stable/2982892.

Robert, Christian, and George Casella. “A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data.” *Statistical Science* 26, no. 1 (2011): 102–15. http://www.jstor.org/stable/23059158.

Roberts, G.O., and A.F. Smith, "Simple conditions for the convergence of the Gibbs sampler and Metropolis–Hastings algorithms.” *Stochastic Process*. Appl. 49 (2) (1993) 207–216.

Rubin, Donald B. “Bayesian Inference for Causal Effects: The Role of Randomization.” *The Annals of Statistics* 6, no. 1 (1978): 34–58. http://www.jstor.org/stable/2958688.

Rosenbluth, M. N., and A. W. Rosenbluth, “Further results on Monte Carlo equations of state,” J. Chem. Phys. 22, 881–884 (1954).

Rosenthal, J. S. “Minorization conditions and convergence rates for Markov chain Monte Carlo.” *J. Amer. Statist. As- soc*. (1995). **90** 558–566. MR1340509.

Shinn, Terry. “Intellectual Cohesion and Organizational Divisions in Science.” *Revue Française de Sociologie* 43 (2002): 99–122. https://doi.org/10.2307/3322759.

Smith, A. “Testing theories of strategic choice: The example of crisis escalation.” *AMERICAN JOURNAL OF POLITICAL SCIENCE* 43, no. 4 (1999): p.1254-1283 ISSN: 0092-5853.

Smith, A. F. M., and G. O. Roberts. “Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods.” *Journal of the Royal Statistical Society. Series B (Methodological)* 55, no. 1 (1993): 3–23. http://www.jstor.org/stable/2346063.

Smith, Michael., and Robert Kohn. “Nonparametric regression using Bayesian variable selection.” *Journal of Econometrics*. Volume 75, Issue 2. (1996). Pages 317-343. ISSN 0304-4076. https://doi.org/10.1016/0304-4076(95)01763-1.

Spiegelhalter D. J. “Bayesian graphical modelling: a case-study in monitoring health outcomes. *Applied Statistics."* 1998; **47**:115–133.

Spiegelhalter, D. J., Andrew Thomas, and Nicky Best. “The BUGS project: Evolution, critique and future directions.” *Stat Med.* 2009 Nov 10;28(25):3049-67. doi: 10.1002/sim.3680. PMID: 19630097.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and W. R. Gilks. “BUGS: Bayesian inference Using Gibbs Sampling, Version 0.5, (version ii).” MRC Biostatistics Unit. (1996a).

Stigler, S. M. “Thomas Bayes's Bayesian Inference.” *Journal of the Royal Statistical Society*. Series A (General) , Vol. 145, No. 2 (1982), pp. 250-258.

Struik, D. J. “On the Sociology of Mathematics.” *Science & Society* 6, no. 1 (1942): 58–70. http://www.jstor.org/stable/40399680.

Tanner, M. A., *Tools for statistical inference : methods for the exploration of posterior distributions and likelihood functions*. New York : Springer-Verlag, c1993.

Tanner, Martin A., and Wing Hung Wong. “The Calculation of Posterior Distributions by Data Augmentation.” *Journal of the American Statistical Association* 82, no. 398 (1987): 528–40. https://doi.org/10.2307/2289457.

Thompson, E. A., and C. A. B. Smith. “Monte Carlo Likelihood in the Genetic Mapping of Complex Traits [and Discussion].” *Philosophical Transactions: Biological Sciences* 344, no. 1310 (1994): 345–51. http://www.jstor.org/stable/56106.

Tierney, Luke. “Markov Chains for Exploring Posterior Distributions.” *The Annals of Statistics* 22, no. 4 (1994): 1701–28. http://www.jstor.org/stable/2242477.

Tierney, Luke. “[Practical Markov Chain Monte Carlo]: Comment.” *Statistical Science* 7, no. 4 (1992): 499–501. http://www.jstor.org/stable/2246103.

University of Cambridge, School of Clinical Medicine, MRC Biostatistics Unit / Software / The BUGS Project at https://www.mrc-bsu.cam.ac.uk/software/bugs/.

University of Cambridge, School of Clinical Medicine, MRC Biostatistics Unit, Software, The BUGS Project, OpenBUGS at https://www.mrc-bsu.cam.ac.uk/software/bugs/openbugs/

University of Cambridge, School of Clinical Medicine, MRC Biostatistics Unit, Software, The BUGS Project, GeoBUGS at https://www.mrc-bsu.cam.ac.uk/software/bugs/thebugs-project-geobugs/

Vleck, T.V. The IBS 7094 and CTSS. (2022). https://multicians.org/thvv/7094.html.

Vounatsou, P., and A. F. M. Smith. “Bayesian Analysis of Ring-Recovery Data Via Markov Chain Monte Carlo Simulation.” *Biometrics* 51, no. 2 (1995): 687–708.

Winkler, Gerhard. *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: A Mathematical Introduction*. Berlin ; New York: Springer-Verlag, 1995.

Wood, W. W., and F. R. Parker. 'Monte Carlo equation of state of molecules interacting with Lennard-Jones potential.' *J. Chem. Phys.* **27,** (1957), 720–733.

Zilsel, Edgar. "The Sociological Roots of Science." *Social Studies of Science* 30, no. 6 (2000): 935–49. http://www.jstor.org/stable/285793.

Appendix A: Metropolis-Hastings algorithm and Gibbs sampling

The steps of the Metropolis-Hastings algorithm are as follows. Consider the case where we want to sample from a distribution $\pi(x)$ but are unable (or very hard) to do so directly. We choose a starting sample x_0 as the initial state and apply the Metropolis-Hastings algorithm. This initial sample can be chosen randomly within the range of the distribution, but a well chosen start will shorten the burn-phases of the chain. Detailed techniques see P13, P135 in Gilks et al. (1996). Then, create a candidate state x' based on $g(x'|x_0)$, which is the likelihood of x' given x_0. $g(x'|x_0)$ depends on the prior knowledge of the different states, and it is sometimes chosen subjectively. A well-designed $g(x'|x_0)$ is essential for the quick convergence of the Markov chain. For more detailed explanation, refer readers to Chib (1995). After the above steps, accept the candidate state with probability $min(1, \frac{\pi(x')}{\pi(x_0)} * \frac{g(x_0|x')}{g(x'|x_0)})$. Here, $\frac{\pi(x')}{\pi(x_0)} * \frac{g(x_0|x')}{g(x'|x_0)}$ measures the likelihood of jumping from the current state to the next state in the Markov chain. It it exceeds 1, then directly accepts x' as x_1; if the kernel is less than one, then accept x' with the probability $\frac{\pi(x')}{\pi(x_0)} * \frac{g(x_0|x')}{g(x'|x_0)}$. In practice, we can generate a random number from a uniform distribution on [0,1], and accept x' if the number is less than $\frac{\pi(x')}{\pi(x_0)} * \frac{g(x_0|x')}{g(x'|x_0)}$. Detailed explanation is in Chib and Greenberg (1995).

If x' is accepted, then set x_1 the same as x'; otherwise set x_1 the same as x_0(the Markov chain does not jump to the next stage). Then, propose a new candidate state x' according to $g(x'|x_1)$, and repeat the process. The Markovian matrix of the chain that we sample from will converge if this strategy is used repeatedly, allowing us to obtain samples from the desired density. We should wait until the algorithm has passed the burn-in states because it takes some time for the Markov chain to reach the stationary distribution, and many of the starting samples might not follow the desired density. The basic steps of the M-H algorithm are listed above. Theoretical justifications of the method are in Hasting (1970).

The other frequently employed MCMC algorithm, Gibbs sampling, samples from the full conditional distributions and is very useful in higher dimensional situations. The updating scheme goes like this: consider a scenario in which there are k variables, $U_1, U_2, \ldots, U_k$. Start by using a baseline set of values $U_1^{(0)}, U_2^{(0)}, \ldots, U_k^{(0)}$. Draw $U_1^{(1)}$ from the conditional probability distribution $[U_1 \mid U_2^{(0)}, \ldots, U_k^{(0)}]$, then $U_2^{(1)}$ from $[U_2 \mid U_1^{(0)}, U_3^{(0)}, \ldots, U_k^{(0)}]$, and so on, all the way up to $U_k^{(1)} \sim [U_k \mid U_1^{(0)}, \ldots, U_{k-1}^{(0)}]$. Every variable (dimension) is guaranteed to be visited in the cycle's prescribed order. At the end of i such iterations, we reach $(U_1^{(i)}, U_2^{(i)}, \ldots, U_{k-1}^{(i)}, U_k^{(i)})$, and have i+1 samples. Once the burn-in period has passed, the Gibbs sampler can offer us samples from the desired high-dimensional density. Theoretical and pragmatical justifications can be found in Geman and Geman (1984), Gelfand and Smith (1990), and Casella and George (1992).

Reference:

Casella, George, and Edward I. George. “Explaining the Gibbs Sampler.” *The American Statistician* 46, no. 3 (1992): 167–74. https://doi.org/10.2307/2685208.

Chib, Siddhartha., and Edward Greenberg. “Understanding the Metropolis-Hastings Algorithm.” *The American Statistician* 49, no. 4 (1995): 327–35. https://doi.org/10.2307/2684568.

Hastings, W. K. “Monte Carlo Sampling Methods Using Markov Chains and Their Applications.” *Biometrika* 57, no. 1 (1970): 97–109. https://doi.org/10.2307/2334940.

Gelfand, Alan E., and Adrian F. M. Smith. “Sampling-Based Approaches to Calculating Marginal Densities.” *Journal of the American Statistical Association* 85, no. 410 (1990): 398–409. https://doi.org/10.2307/2289776.

Geman, S., and D. Geman. “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images.” *IEEE Trans Pattern Anal Mach Intell.* (1984). Jun;6(6):721-41. doi: 10.1109/tpami.1984.4767596. PMID: 22499653.

Gilks, W. R. , S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London. (1996).

Appendix B: The basic Bayesian updating process

The subjective interpretation of probability views the unknown parameters in Bayesian inference as distributional characteristics; they are not treated as constants, but as variables with density functions. Bayesian statisticians can therefore provide subjective distributions of parameters rather than point estimates as hypotheses. The Bayesian inference process can also be summed up in three steps. Initially, specify a probability model with prior information regarding the unknown parameter values. Using Bayes' theorem, update knowledge of the unknown parameters based on the observed data. The outcome is known as the posterior probability. Evaluate the model's compatibility with the data and the sensitivity of its assumptions (Gill 2002).

The form of the Bayes theorem, the essential mathematical tool for Bayesian inference, is as follows: $P(\mathrm{A} \mid \mathrm{B}) = P(\mathrm{A})\frac{P(\mathrm{B} \mid \mathrm{A})}{P(\mathrm{B})}$. $P(A)$ and $P(B)$ denote the probability of A and B, respectively; $P(\mathrm{A} \mid \mathrm{B})$ is the conditional probability of A given that B is true.

Example of the update can be given like this: let A denote the event that a person get covid, and B denotes the person was tested positive. Suppose we know that a covid test is 90% accurate for positive ($P(\text{positive} \mid \text{get covid})$ = 90%), and 99% accurate for the negative($P(\text{get covid} \mid \text{negative})$ = 1 - 99% = 1%). Now we still cannot derive the probability of me getting covid given tested positive ($P(\text{get covid} \mid \text{positive})$). We must have a prior probability $P(\text{get covid})$. Suppose it is 1%. Then, by Bayes' theorem,

$$P(\text{get covid} \mid \text{positive}) = P(\text{get covid})\frac{P(\text{positive} \mid \text{get covid})}{P(\text{positive})}$$

$$= 0.01 * \frac{0.10}{0.01 * 0.90 + 0.99 * 0.01} \approx 0.053$$

, which is minimal compared with the 90% accuracy. In this example, we have an accurate test (99% for the negative and 90% for the positive), but due to the prior knowledge, the probability of getting covid given tested positive is as low as 5.3%. This is a basic application of Bayesian update of our degrees of belief. The significance of the prior probability in statistical inference when the number of steps are limited can be seen in this example.

Bayes' theorem enables us to revise our belief in light of the prior probability distribution of A and the observed data B. The Bayes Theorem can be extended to a multitude of events. Assume we are interested in the probabilities of three exhaustive and mutually exclusive events A, B, and C given the data. D. The Total Probability Law states

$$P(D) = P(D|A) * P(A) + P(D|B) * P(B) + P(D|C) * P(C)$$

, so $P(\mathrm{A}|\mathrm{D})$, the conditional probability of A given D, is

$$\frac{P(\mathrm{D}|\mathrm{A})P(\mathrm{A})}{P(D|A) * P(A) + P(D|B) * P(B) + P(D|C) * P(C)}.$$

Similar formulas is applied for B and C.

The Bayesian formula for updating confidence can be generalized to cases involving multiple and even continuous circumstances. For N mutually exclusive and exhaustive events (hypotheses) $\theta_i, i = 1,2,...,N,$ the probability of certain event θ_i given data D is

$$P(\theta_i|D) = \frac{P(D|\theta_i)P(\theta_i)}{\sum_{j=1}^{N} P(D|\theta_j) * P(\theta_j)}.$$

The formula provides us with the posterior distribution $P(\theta_i | D)$. That is the Bayes theorem-based procedure for updating credibility in a discrete case. In continuous situations, the posterior distribution $\pi(\theta | \text{data})$ is obtained by $\frac{g(\theta) \times f(\text{data} | \theta)}{\int g(\theta) \times f(\text{data} | \theta) d\theta}$, where g and f are density functions representing the probability distribution.

Since the denominator $\int g(\theta) \times f(\text{data} | \theta) d\theta$ is fixed and does not affect the relative probabilities for A, it can be considered a normalizing constant that ensures the distribution's integral is one (P16, Gill 2002).

Thus, the formula can be rewritten as $\pi(\theta | \text{data}) \propto g(\theta) \times f(\text{data} | \theta)$, also known as the Bayesian Mantra: a posterior is proportional to the prior multiplied by the likelihood. This formula provides Bayesian statisticians with an analytical technique for determining the precise posterior distribution. Detailed examples of Bayesian update analysis can be found in (P41, Gelman 1995).

Reference:

Gelman, Andrew. *Bayesian Data Analysis*. London; New York: Chapman & Hall, 1995.

Gill, Jeff. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, Fla.: Chapman & Hall/CRC, 2002.

Appendix C: Frequentist inference

Frequentist inference focuses primarily on point estimation and hypothesis testing. It consists generally of three steps: empirical establishment of stable long-run frequencies of events, guessing and verifying the repeated operations that produce the observed frequencies, and using "hypothetical chance mechanism" to deduce the rules for adjusting our decisions (Neyman 1977).

Consider the following example: imagine we have a hypothesis regarding the mean height of students in Chicago. On the basis of the characteristics of a sample, frequentist inference can help us obtain a point estimate — a precise number as the guess — based on a point estimate (mean, variance, etc.). We may also construct a confidence interval based on the criteria we select (for instance, 95 percent). Even when there is an interval, this does not indicate that the true parameter has a distribution, as frequentists always see the parameter as an unknown value and not a variable. Consequently, the right interpretation of a 95% confidence interval is that 95% of hypothetical intervals would include the true value if the study were duplicated numerous times (Bijak, 2016).

The objective interpretation leads to this easier (compared with Bayesian) approach. The fundamental idea is to treat the parameters as objective and fixed values, and to estimate and approximate their true values using data. Then, we can conduct hypothesis test. The steps of a Fisherian null hypothesis test are as follows: Establish the null hypothesis Under the assumption that the null hypothesis is true, determine the test statistic and its distribution. Determine the test statistic from the collected data. Calculate the significance level corresponding to the test statistic under the null hypothesis using the distribution. Determine whether to reject the null hypothesis based on the level of significance. Then, Neyman and Pearson advocated a similar, but more rigorous approach: propose two complimentary hypotheses together, and analyze, or test on based on the other. Such a method provides further information about the different types of errors, power, etc. (Gill 1999). Consequently, in the Fisher-Neyman-Pearson method, the p-value, the confidence interval, and the significance are essential notions, which has been criticized by Bayesian statisticians.

Reference:

Bijak, J. "Bayesian demography 250 years after Bayes." *Population Studies.* Vol. 70, No. 1 (MARCH 2016), pp. 1-19.

Gill, Jeff. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52, no. 3 (1999): 647–74. https://doi.org/10.2307/449153.

Neyman, J. "Frequentist Probability and Frequentist Statistics." *Synthese* 36, no. 1 (1977): 97–131. http://www.jstor.org/stable/20115217.

Figures

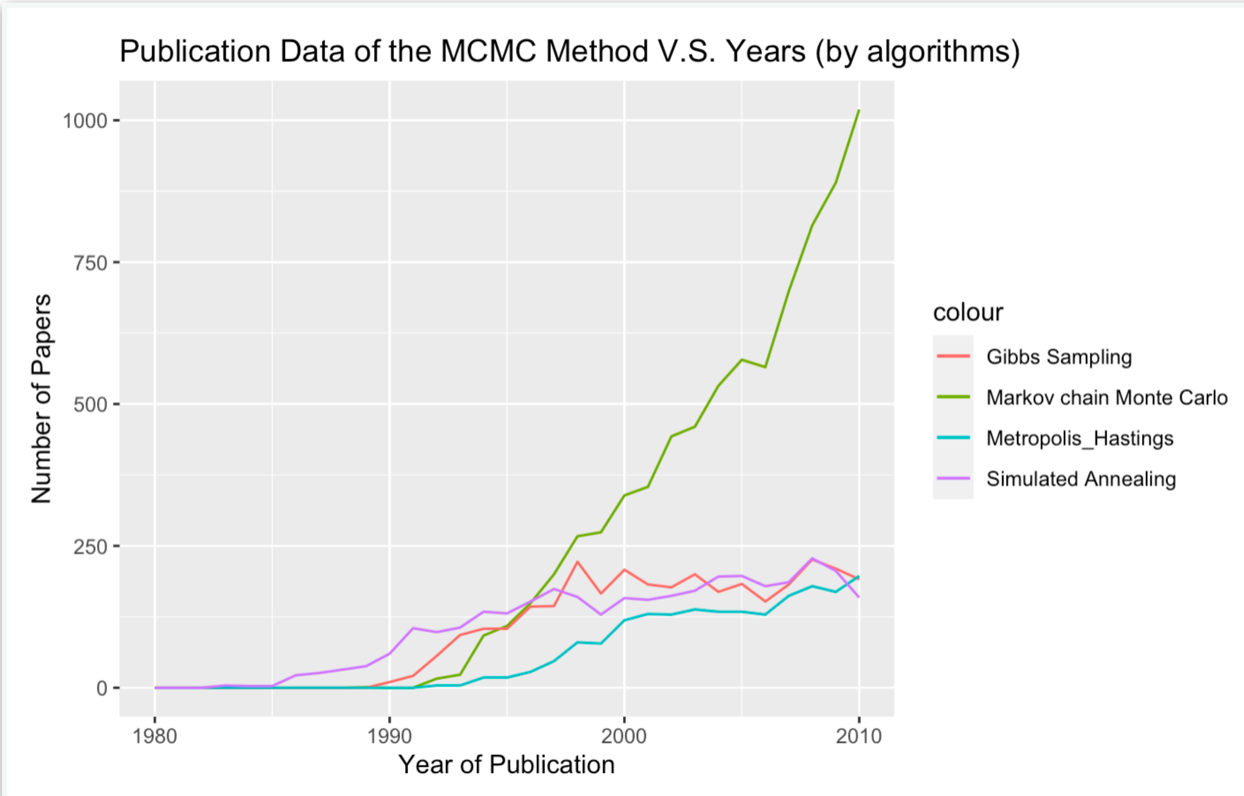

Figure 1

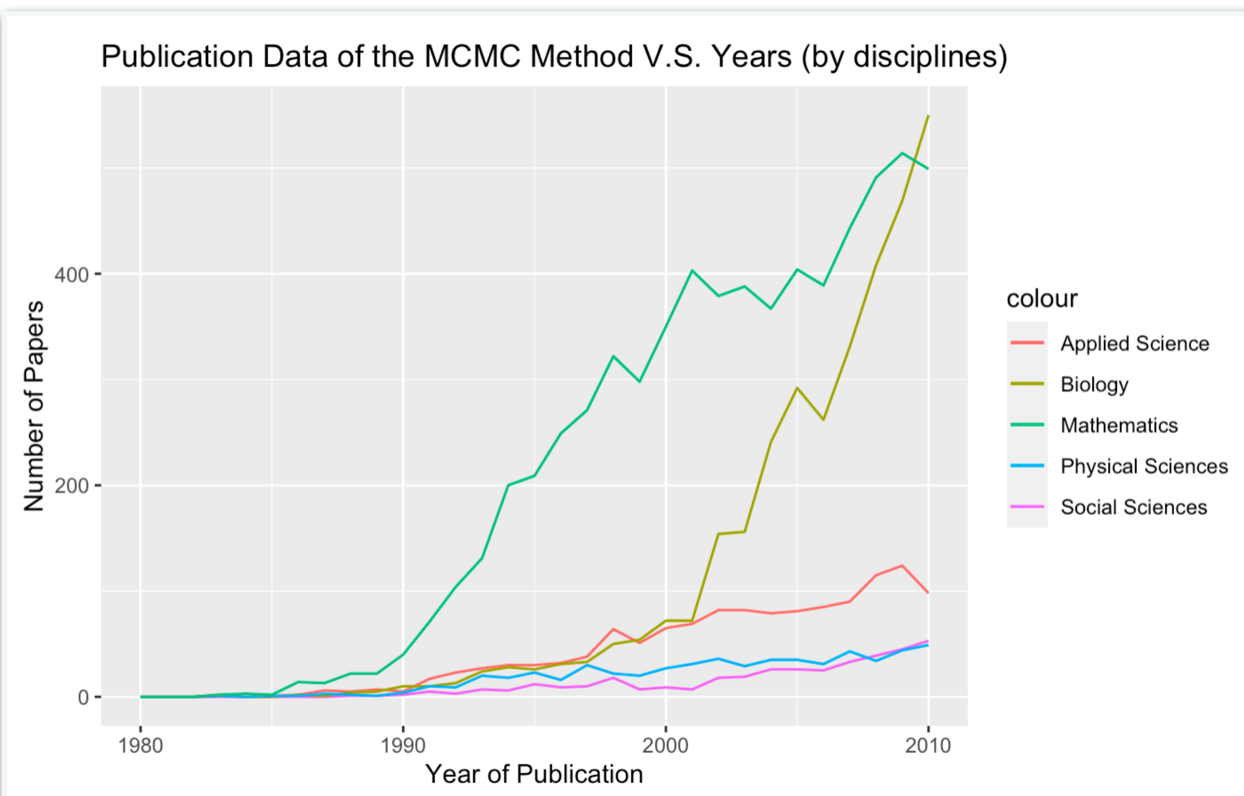

Figure 2

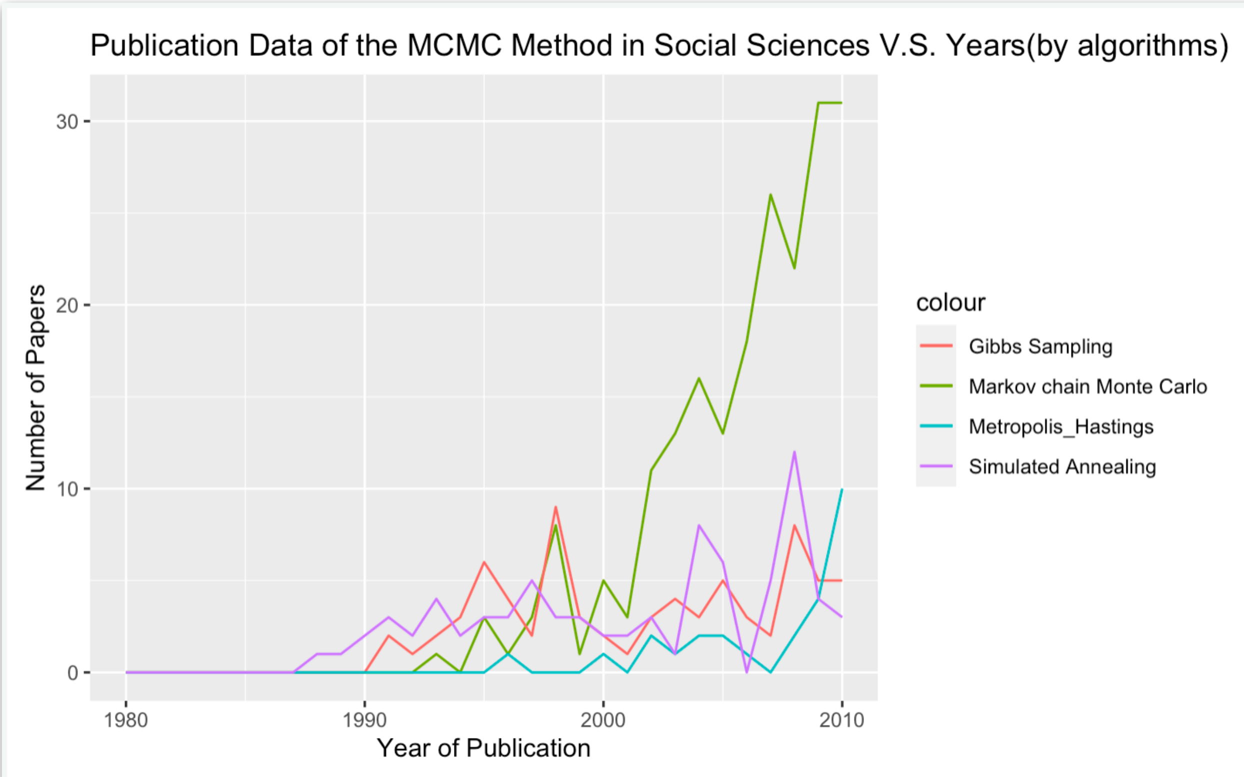

Figure 3

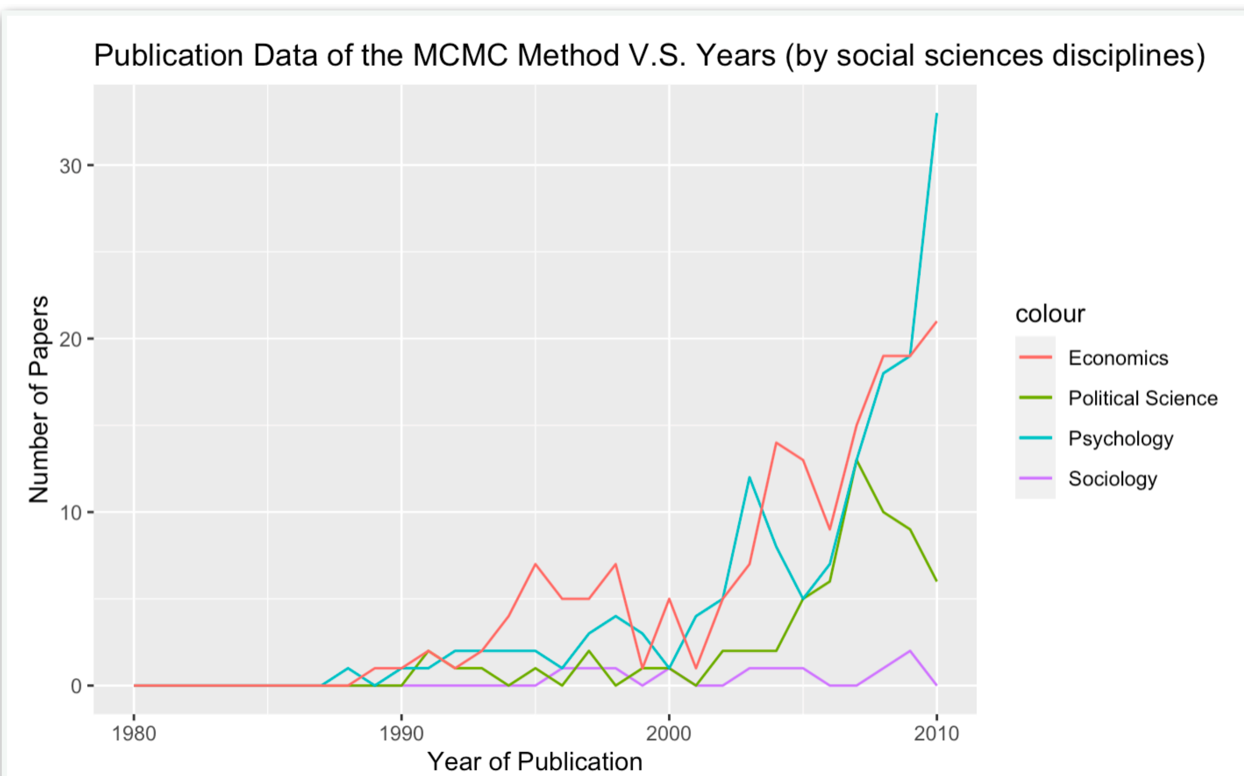

Figure 4