A Transformation of Bayesian Statistics:
Computation, Prediction, and Rationality

Johannes Lenhard, RPTU, Kaiserslautern, Germany

To appear in: Studies in History and Philosophy of Science 92 (2022) 144–151

Abstract

Bayesian approaches have long been a small minority group in scientific practice, but quickly acquired a high level of popularity since the 1990s. This paper shall describe and analyze this turn. I argue that the success of Bayesian approaches hinges on computational methods that make a class of models predictive that would otherwise lack practical relevance. Philosophically, however, this orientation toward prediction comes at a price. The new computational approaches change Bayesian rationality in an important way. Namely, they undercut the interpretation of priors, turning them from an expression of beliefs held prior to new evidence into an adjustable parameter that can be manipulated flexibly by computational machinery. Thus, in the case of Bayes, one can see a coevolution of computing technology, an exploratory–iterative mode of prediction, and the conception of rationality.

Keywords: Bayes, computational modeling, Markov chain Monte Carlo, philosophy of statistics, prediction, rationality, scientific practice

1. Introduction

If statistics is viewed as a branch of mathematics, it has to be seen as a special branch distinguished by the ways in which it is linked to societal practices and to philosophical positions.[1] Bayesian statistics is exemplary on both counts. Philosophers have discussed Bayesian

[1] The history of probability casts light on how science and society, in mutual interrelation, developed ways to deal with uncertainties ranging from strategies in games to the prices of annuities. Whereas predicting the death of an individual might require divine foresight, estimating the death rate in a large population

statistics vigorously and elaborated Bayesianism as a *philosophical* position.[2] Most significantly, Bayesian epistemology analyzes how one should deal with new data in a rational way—that is, the Bayesian standpoint lays claim to capture scientific rationality. Put simply, Bayes' rule[3] is taken as a (normative) principle that prescribes how one should update prior beliefs in the light of new evidence. The use of Bayesian approaches in scientific *practice* shows a remarkable career. Despite their philosophical prominence, they remained a small minority group in science—but only up to the 1990s when Bayesian methods quickly acquired a high level of popularity in the sciences as well.

This paper shall describe and analyze this turn. I argue that the success of Bayesian approaches hinges on computational methods that make a class of models predictive that would otherwise lack practical relevance. Philosophically, however, this orientation toward prediction comes at a price. The new computational approaches change Bayesian rationality in an important way. Namely, they undercut the interpretation of priors, turning them from an expression of beliefs held prior to new evidence into an adjustable parameter that can be manipulated flexibly by computational machinery—a lubricant for exploratory iteration. Thus, in the case of Bayes, one can see a coevolution of computing technology, an exploratory–iterative mode of prediction, and the conception of rationality.

Section 2 briefly introduces the rift in the philosophy and practice of statistics in which the Bayesian and the classical accounts were used, elaborated, and defended by different fields and disciplines—until the popularity of Bayesian methods unfolded in the 1990s. According to the prevailing stance in philosophy, the advantages in terms of rationality account for the upswing of Bayesian methods (section 3). Contrary to this view, I claim that it was the move to an iterative–exploratory mode of prediction—on the technological base of cheap and easily available computers—that drove this upswing. This claim is supported with an analysis of the

---

can get by with profane data. The historical and philosophical accounts by Daston (1988), Hacking (1990), or Porter (1995) cover the seventeenth to nineteenth centuries in admirably sophisticated ways.

[2] The *Stanford Encyclopedia of Philosophy* has entries on the philosophy of statistics (Romeijn, 2017) and a separate one on Bayesian epistemology (Talbott, 2016). Taken together, these provide a guide to the large body of philosophical literature on Bayesianism.

[3] This rule follows from the definition of conditional probabilities and is accepted unquestionably across all camps.

pivotal roles played by Markov chain Monte Carlo methods (section 4) together with software packages (section 5). The concluding section 6 gives an outlook on the pragmatic stance that has gained ground in philosophy of statistics over the last two decades.

2. Bayes' Popularity

Before appreciating its growing popularity, I shall briefly and, as specialists will rightly bemoan, superficially describe the situation before it started. Bayes' rule captures how to calculate with conditional probabilities. Let $\pi(H)$ stand for the probability of a statement or hypothesis H, and $\pi(H \mid D)$ for the conditional probability of H given D. Now, both H and D happen if (for the moment, think of a temporal order) D happens and then H happens given D, or equivalently, H happens and then D happens given H. In other signs: $\pi(D) \cdot \pi(H \mid D) = \pi(H) \cdot \pi(D \mid H)$. Separating $\pi(H \mid D)$ on the left side gives Bayes' rule:

$$(*) \qquad \pi(H \mid D) = \pi(H) \cdot \pi(D \mid H) / \pi(D).$$

It is named after Reverend Thomas Bayes (c. 1701–1761), a Presbyterian minister, philosopher, and statistician. Bayesianism starts out with a special interpretation of this rule. Consider you have some hypothesis H—for example, that it will rain tomorrow. You do not know for sure, so (in a Bayesian mood) the degree of your belief can be expressed as a probability, $\pi(H)$. Now there arrives new evidence D—say, you stand up next morning and have a look at the sky. This should give you additional evidence and will change your (subjective) probability of rain on this day. Therefore, $\pi(H)$ is also called the "prior" that will be updated. The updated probability, written $\pi_{D(H)}$, of your hypothesis given the data is also called the "posterior." Which numerical value does it have? Bayesians take the position that updating needs to happen by conditionalization. The posterior is the conditional probability: $\pi_{D(H)} = \pi(H \mid D)$. In other words, equation (*) answers the question: The posterior is proportional to the (subjective) prior $\pi(H)$ and to $\pi(D \mid H)$, the so-called likelihood—that is, the probability of the data given your hypothesis (how likely the sky looks like it does in the morning given that it will rain). The term $\pi(D)$ plays the role of a (normalizing) constant.

Bayesianism adopts (*), or a sophisticated variant of it, as a *principle* that *should* guide inferences. In the entry to the *Stanford Encyclopedia*, to present a generic point of view,[4] W. Talbott (2008) identifies the main features of Bayesian epistemology as the introduction of a formal apparatus for inductive logic that uses the laws of probability as coherence constraints on rational degrees of belief. In particular, it takes Bayes' rule (a basic rule for conditional probabilities) as a norm for probabilistic inference, as a *principle of conditionalization*. "What unifies Bayesian epistemology is a conviction that conditionalizing . . . is rationally required in some important contexts—that is, that some sort of conditionalization principle is an important principle governing rational changes in degrees of belief." Famous arguments from Bayesian epistemology, such as the Dutch book argument, set out to show that following Bayes' principle is following a demand of rationality.[5]

The classical camp of, among others, Fisher, Neyman, and Pearson—despite internal differences[6]—criticized mainly two points: First, Bayesian estimations hinge on subjective priors and are therefore not robust. Any robust results would have to take into account the variability of priors—that is, other probability measures that do not correspond to the actual beliefs.[7] Statistical inference should be geared toward the properties of the estimation (such as robustness) rather than rationality according to a system of beliefs. Second, the Bayesian assumptions create high obstacles for practice. Calculating with **(*)** does not only require the specification of all probabilities involved: the probability of a hypothesis $\pi(H)$, the probability of the data $\pi(D)$ (often expressed via conditioning on different possibilities), and the conditional probability $\pi(D \mid H)$. Crucially, their numerical values have to be computed. In a technical sense, the calculation of

---

[4] A sample of standard accounts from both sides of the Bayesian versus classical divide is Earman (1992), Howson and Urbach (1993), and Mayo and Spanos (2009). Hacking (2001) provides an accessible introduction. I openly admit that the picture I paint does not match the complexity of extant terminology in which subjectivism is pitted against objectivism, frequentist accounts of probability against subjective ones, and so on.

[5] Trying to keep things simple, I have glossed over internal differentiations of Bayesians. Neal's (1998) verdict that "there is (in theory) just one correct prior" might be controversial among adherents of Bayesianism. Corfield and Williamson (2001), for instance, discern subjective priors from objective (pluralist, logical, empirical) ones that, however, add further requirements for being rational while maintaining the principle.

[6] On these differences, see, for instance, Lenhard (2006).

[7] Bayes, for instance, had presented an example in which he did not know about priors and assumed equal distribution among possibilities. Neyman considered this step illegitimate.

posterior probabilities requires an evaluation of very difficult integrals. It is preferable, according to the classical camp, to avoid specification and computation of this kind. Classical statistical modeling aimed to do without priors—such as the famous "null hypothesis" in significance tests that allows researchers to be agnostic.[8]

The rift between the Bayesian and classical camps was reflected by a divide of disciplines. While economics, and—of course—philosophy have been (and still are) dominated by Bayesianism, in most natural sciences, it is classical methods that have a stronger footing, although this is certainly not a clean divide. This brought philosophy of science into an odd position. Philosophers worked out a Bayesian normative account, whereas large parts of the sciences apparently did not care, but rather continued to prefer classical approaches. Here is a typical opinion from a Bayesian statistician reasoning about why the uptake in scientific practice was so slow.

> Bayesians were still a small and beleaguered band of a hundred or more in the early 1980s. Computations took forever, so most researchers were still limited to "toy" problems and trivialities. Models were not complex enough. The title of a meeting held in 1982, "Practical Bayesian Statistics," was a laughable oxymoron. One of Lindley's students, A. Philip Dawid of University College London, organized the session but admitted that "Bayesian computation of any complexity was still essentially impossible (…). Whatever its philosophical credentials, a common and valid criticism of Bayesianism in those days was its sheer impracticability." (McGrayne, 2011, pp. 213–214, quote from interview)[9]

However, the divide changed in a remarkably swift way. Figure 1 presents some bibliometric evidence. The data are from the *Web of Science* and count papers appearing in one of five major statistics journals: *The Journal of the Royal Statistical Society B*, *Annals of Statistics, Journal of the American Statistical Society, Biometrika*, and *Biometrics*. Each point shows the percentage of papers (in one particular year) whose topic contains "Bayes." All five journals come from the classical side that dominated mathematical statistics. The data confirm the outsider role of

[8] Bayesians, in turn, would typically object that such agnosticism ignores relevant knowledge that actually *is* available.

[9] McGrayne's book is about the eventual success of Bayesian approaches, so the quote does not reflect a bias against Bayesianism.

Bayesian methods in professional scientific statistics, with a consistent share of only 2–4% up to the early 1990s. Then, however, there is a rapid rise to a level of about 20%.[10]

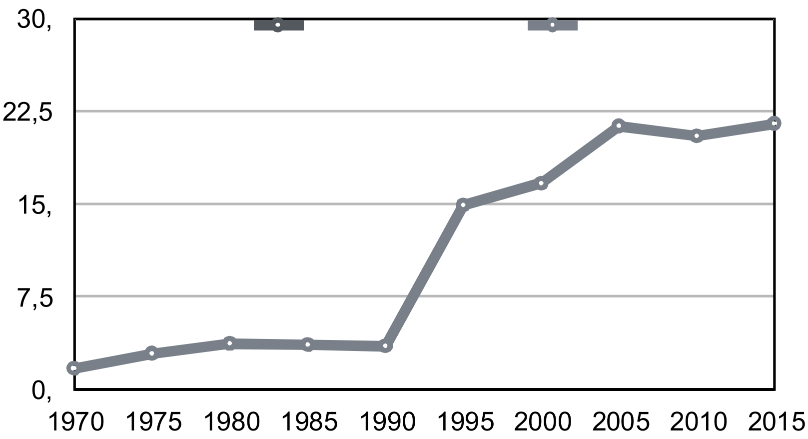

Figure 1. Axes are time (years) and percentage. Ten data points are displayed, one every fifth year, 1970 to 2015. Straight lines are added connecting these points.

This picture is conservative, because it displays (at least three) very traditional journals that will surely not over-represent papers inclined toward Bayes. Newly established journals tend to show an even higher share, but cannot provide reference points for the mid-twentieth century.[11]

Bayesian methods eked out an existence as a small minority group in the sciences and their statistical approaches—up to the early 1990s. After that, Bayesian methods developed quickly, indeed almost leapt up to become an intensely researched and widely used approach. Since then, the extent of literature on Bayesian methods in the sciences has grown rapidly. One can track this in many forms from journal papers and discussion statements to books and

---

[10] This figure resembles the one included in chapter 4 on the rise of density functional methods. Both depict a 1990s turn.

[11] There are also some areas of statistical work that are closely connected to Bayesian methodology. One example would be causal analysis and Bayesian nets, a field following the lead of Judea Pearl (cf. Pearl, 1995; or Williamson, 2005). At present, it is a subfield of artificial intelligence and also philosophy of science. The flourishing of examples of this sort is not included in Figure 1 that is already dramatic enough to motivate my analysis.

encyclopedia articles. The (Bayesian inclined) statisticians Bradley Carlin and Thomas Louis (2000), for instance, note:

> An impressive expansion in the number of Bayesian journal articles, conference presentations, courses, seminars, and software packages has occurred in the four years since 1996 . . . Perhaps more importantly, Bayesian methods now find application in a large and expanding number of areas where just a short time ago their implementation and routine use would have seemed a laudable yet absurdly unrealistic goal. (p. xiii)

Thus, I take it for granted that the uptake and use of Bayesian methods experienced a turn in the 1990s, and we devote the remaining sections to analyzing this turn.

3. Rationality or computation?

The turn did not go unnoticed from the side of either Bayesian statisticians or philosophers. We shall complement the Carlin and Louis quote above with one taken from the philosophical side. In their volume on the foundations of Bayesianism, Corfield and Williamson (2001, p. 3) offer an outlook on the field: “Bayesianism has emerged from being thought of as a somewhat radical methodology—for enthusiasts rather than for research scientists—into a widely applied, practical discipline well-integrated into many of the sciences.” Scientists and philosophers agree unanimously that the turn happened. The next question is: Why did it happen?

A common viewpoint holds that the main reason for the turn is the rationality of Bayesianism itself that finally became operational thanks to computational methods. Computer-based methods rendered feasible the integrations (e.g., when calculating conditional probabilities in complex models) that Bayes’ rule requires; and, as a result, the rule’s rationality gained traction.[12] As Corfield and Williamson (2001) put it (looking back on the 1990s), it is only recently that “computers have become powerful enough, and the algorithms efficient enough, to perform the integrations” (p. 4). Although this explanation is plausible, it is crucially incomplete and can therefore easily mislead.

The point is that simply using computational tools does not lead straightforwardly to obtaining those results that had been too difficult to achieve before. The tools are not strictly

[12] Zellner (1988) and Howson and Urbach (1993, 1st ed. 1989) argue for the superior rationality of Bayesianism independently of computational methods. Hence, they cannot account for the timeline of the upswing (except that it had to happen sometime).

neutral. Which mathematical tools are used and how they are used might influence the modeling process. This is exactly my key point. The analysis shows that, thanks to new computational tools, Bayesian methods changed into a new, exploratory–iterative mode of prediction. Furthermore, I argue, this mode of prediction affects the very nature of Bayesian rationality.

I am well aware that this claim is not easy to substantiate. It ascribes a significance to computational methods that has not always been apparent to practitioners. Early appraisals of the computer and its powers typically took it for granted that the machine would simply carry out logical or arithmetical operations and would not require any new perspective on how mathematical tools lead to predictions. For instance, the statistician Dennis Lindley, a leading advocate of Bayesianism over almost the entire second half of the twentieth century, had seen Bayes' rule as an arithmetic recipe for producing inferences. He considered this procedure to be almost mechanical, given that the integrations could be made feasible (Lindley, 1965) Lindley did not see any particular interest in devising computational methods. The next generation of Bayesian-minded statisticians, however, saw things differently. The statistician A. F. M. Smith, a leading voice, argued in a sort of manifesto that it was efficient numerical integration procedures that led to the more widespread use of Bayesian methods (Smith, 1984).

Even granted the importance of numerical procedures, it was hard to anticipate just how such procedures would change the method. Identifying the computational tools on which the Bayesian boom is built is straightforward. There is ample evidence in which statisticians write about what created the difference in the 1990s. In fact, there is a remarkable consensus on this point: It was the Markov chain Monte Carlo (MCMC) method that made the difference—and Smith himself provided a key paper.

> When Smith spoke at a workshop in Quebec in June 1989, he showed that Markov chain Monte Carlo could be applied to almost any statistical problem. It was a revelation. Bayesians went into "shock induced by the sheer breadth of the method." By replacing integration with Markov chains, they could finally, after 250 years, calculate realistic priors and likelihood functions and do the difficult calculations needed to get posterior probabilities. (McGrayne, 2011, pp. 221–222)[13]

[13] Similar quotes abound. We picked this quote from McGrayne's book because of its atmospheric qualities. Here are two alternative quotes: "In fact, it may be argued that the main reason that the Bayesian

Thus, MCMC opened the door for Bayes to become practically relevant. There is agreement on this point. Now is where the difficult part begins. I shall analyze how MCMC affects the very rationale of Bayesianism.

4. The Markov chain Monte Carlo revolution: Iteration and exploration

This section bears the main thrust of the argument. Using Markov chain Monte Carlo (MCMC) methods, I argue, yields predictions in an iterative–exploratory mode and thus affects the rationale of Bayesian methods.

integration and convergence

I start with a brief summary of what MCMC is about. Although this part will be about integration in a technical mathematical sense, I shall keep it on a nonformal level. From the outset, MCMC combines Monte Carlo, which stands for iteratively sounding out mathematical terms, with Markov chains, a class of random processes.

Monte Carlo strategies are based on the law of large numbers. This law states that the expected value of a random variable is approximated by the average of many random trials (that each follow the same probability distribution). In a casino such as the one in Monte Carlo, many people gather around a roulette table after a small series of identical outcomes happens—say three times "13." Some think that the next trial is likely to be the number 13 again; others hold that a different number has to come now. Nonetheless, to the extent that the owner of the casino lets the roulette wheel operate as a (near perfect) generator of random outcomes, the law of large numbers will defy any superstitious beliefs and apply relentlessly: In the long run, the number 13 will make up 1/37 of all numbers (0–36).

---

approach to statistics has gained ground compared to classical (frequentist) statistics is that MCMC methods have provided the computational tool that makes the approach feasible in practice" (Häggström, 2002, p. 47). The probability theorist here agrees with statisticians: "A principal reason for the ongoing expansion in the Bayes and EB [empirical Bayes, jl] statistical presence is of course the corresponding expansion in readily-available computing power, and the simultaneous development in Markov chain Monte Carlo (MCMC) methods and software for harnessing it" (Carlin and Louis, 2000, p. xiii).

It is the simplicity of the example that makes Monte Carlo look trivial. Think of another example: A friend gives you a map of Norway—a country with a famously fractal-like coastline[14]—and asks how large is the area of Norway. In mathematical theory, integration provides the answer; in practice, however, integration can be carried out only for a narrow (and relatively simple) class of functional descriptions. Monte Carlo can help out—just hang the map on the wall, paint a big square around it, and throw darts at the wall. The number of darts that hit the map relative to the number of darts that hit the square approximates the area of the board relative to the square on the wall, which is easy to measure—just count. Of course, this finding hinges on two conditions,[15] namely, that many darts are thrown and that they are distributed randomly across the entire square. For humans, it is hard to fulfill these conditions. But they are almost tailor-made for a computer. One can readily simulate the random procedure and iterate it millions of times—and thereby approximate the integral.[16] Although Monte Carlo is almost tailor-made for the computer insofar as it transforms a problem of *integration* (an operation of calculus) into a problem of *iteration*, it is not immune to the curse of complexity. The simulated value—that is, the fraction of hits among all trials, converges only slowly toward the (unknown) integral. Even after a large number of iterations, the simulated value might not be very accurate, so that, despite the high speed of modern computers, Monte Carlo is, in many instances, ineffective.

This is where the second component of MCMC comes into play. Markov chains are processes that move in a space according to rules of a certain type. For every state or location[17] in this space, there is a list of what the possible locations are that the process can reach with the next step, plus a probability distribution according to which the next-step location from this list will be chosen. In other words, the next step of the process depends only on its present location (and the random choice to be made in this step) and not on the history of the process. One can imagine that at each location, the rule for where to possibly move is written on a signpost—such as “with probability x go one step north, with probability 1-x jump hundred steps south—also called the

---

[14] For a short discussion on the form, beauty, and creation of Norway, see Adams (1979).

[15] A third condition—that this should be carried out in your friend’s apartment rather than your own—is of merely aesthetical concern and mathematically irrelevant.

[16] The term “evaluation of integral” more aptly describes the point than “numerical integration.”

[17] We use the notion of location that suggests a geographical picture. Using the term state would be less intuitive, but more apt to the generality of MCMC methods.

transition probability for each location. The rules might be complicated, but they never refer to where you come from (never state something like “if you are here for the third time, do . . .”). In other words, moving according to these rules does not require memory: Just execute the rules written on the signpost where you stand. Markov chains are often called random processes without memory.

The basic theorem about Markov chains states that such a chain will converge to its stationary (equilibrium) distribution no matter where it started.[18] In other words, in the long run, the process will visit each location in a certain fraction of all steps. Some locations are visited more often; others, only rarely—reflecting the equilibrium distribution. The astonishing and crucial observation is that this convergence happens very quickly. MCMC is based on this observation. The pieces come together for numerical integration in the following way: First, there is an unknown integral one can describe but not evaluate, such as the posteriori probabilities in Bayes’ rule. Assume one can refashion this integral as the stationary distribution of a Markov chain. Then the recipe is straightforward: Simulate the Markov chain for many steps (easy iteration for computers) until it is in equilibrium and record its value. Reiterate this many times (Monte Carlo). The average over all values obtained then approximates the (unknown) integral. The trick depends on two conditions: First, one must find a way to interpret an unknown integral as an (unknown) equilibrium distribution of a Markov chain. Second, the Markov chain must have reached its stationary distribution before one samples its value. The first condition sounds more difficult than it actually is, whereas the second condition sounds easy but is not. I shall discuss both conditions in turn.

application: the MCMC trick

The MCMC method was invented early on in the pioneering times following the creation of the digital computer. It goes back to the work of Nicholas Metropolis, Stanislaw Ulam, and others at Los Alamos and received a classical generalization by Hastings (1970). However, it took another twenty years before MCMC started to take off when examples became available that showed how powerful and flexible the method is—in particular, how doable it is to refashion

[18] We greatly simplify matters in this discussion. Questions regarding how the space is defined or which technical conditions have to be satisfied are not important for the illustrative task we are pursuing here.

complicated integrals as stationary distributions of Markov chains. One famous instance is the Ising model that describes how spins (up or down) on a grid interact with their neighbors. The model is famous not only because the simple interaction can lead to phase transitions and other surprising behavior but also because the problem of determining its equilibrium proved to be utterly unsolvable by analytical means and had become a mathematical monument of intractability. It turned out that MCMC could approximate this distribution with a surprisingly moderate effort in modeling as well as computation.[19]

The Ising model is not a singular case. Mathematicians and statisticians quickly realized that the wide applicability of MCMC to long-standing problems of integration changed the game regarding computational tractability. Restrictions to the mathematically convenient could be lowered substantially, and "from now on, we can compare our data with the model that we actually want to use . . . This is surely a revolution" (Clifford, 1993, p. 53). Many actors agree with seeing this as a revolution. Diaconis (2009), for instance, provides an insightful treatise on "The Markov Chain Monte Carlo Revolution."[20] Part of his treatise is worries about the speed of convergence (our second condition) that we shall discuss below. Put plainly, the revolution consisted in how far the limits of mathematically—and statistically—tractable models have been extended.

On the side of the practitioner, the main benefit is flexibility in modeling. Bayes' rule became practical for a wide array of models. Although it required the evaluation of posteriors, thanks to MCMC, they lost their horror. A wide array of Bayesian applications followed the availability of MCMC; computational approaches in fields such as statistical physics, molecular simulation, bioinformatics, or dynamic system analysis started to flourish. Statistician Jeff Gill called the combination of Bayesianism and MCMC "arguably the most powerful mechanism ever created for processing data and knowledge" (Gill, 2002, p. 332).[21]

---

[19] Persi Diaconis, a leading expert on MCMC, describes the jaw-dropping surprise when he first saw how MCMC solved this task (2009). R. I. G. Hughes (1999) gives a good account of the Ising model in the context of modeling and simulation.

[20] Titterington (2004, p. 192) makes a case about a "Bayesian computational revolution"; Smith and Roberts (1993, p. 4) make a similar case.

[21] MCMC is not necessarily Bayesian, but protagonists of Bayesian approaches were often developing MCMC methods to make Bayes' rule more relevant to practice.

One prominent example from the growing set of MCMC variants is the "Gibbs sampler" for treating inferences involving images with many pixels.[22] It was invented by the Geman brothers as a variant of Monte Carlo and gained enormous traction when Gelfand turned it into a MCMC method.[23]

> The trick was to look at simple distributions one at a time but never look at the whole. The value of each one depended only on the preceding value [The Markov "no memory" property, jl]. Break the problem into tiny pieces that are easy to solve and then do millions of iterations. So, you replace one high-dimensional draw with lots of low-dimensional draws that are easy. The technology was already in place. That's how you break the curse of high-dimensionality. (Gelfand, quote from interview, McGrayne, 2011, p. 221)

Thus, the Gibbs sampler construes a Markov process moving through simple distributions. Thanks to his inventive imagination, Gelfand saw how an intractable object (a high-dimensional distribution) arises from much simpler objects (a process moving through simple distributions). Much like the equilibrium distribution of the Ising model is built from a process that moves through simple distributions (simple flips of one spin). The MCMC trick replaces a computationally intractable object by very many iterations of simpler, tractable objects.[24]

exploration and flexibility

MCMC has an iterative nature. It also has an exploratory nature. When proponents such as Smith and Roberts (1993) state that MCMC methods are for "exploring and summarizing posterior distributions in Bayesian statistics" (p. 3), the point about exploration is important. Exploration plays a role on two different levels. Firstly, modeling approaches quite generally explore, including Bayesian statistics in particular: You always explore what the data are telling

---

[22] Work on the Gibbs sampler started with Geman and Geman (1984) and gained popularity rapidly after the landmark paper of Gelfand and Smith (1990). The hit-and-run algorithm is related to the older Metropolis–Hastings and was proposed by Bélisle et al. (1993) and Chen and Schmeiser (1996). Other MCMC methods show similar timelines.

[23] McGrayne tells the story in a vivid way on pp. 218 ff. "The minute Gelfand saw the Gemans' paper, the pieces came together: Bayes, Gibbs sampling, Markov chain, and iterations" (p. 221).

[24] Philosophers of science have argued that computer simulation changes mathematics because simulation enlarges the realm of tractability (see, e.g., Humphreys 2004). Regarding iteration, this is certainly correct. From this perspective, MCMC is a way to utilize the new tractability for problems of modeling.

you relative to a model that you confront with those data.[25] Exploration of this sort is at the heart of modeling—in lack of complete knowledge one explores with the help of models. Exploration also happens on a second level, exploring the mathematical model itself. And this is where MCMC becomes relevant.

MCMC methods *simulate* relevant properties of mathematical objects (such as integrals or distributions) in numerous iterated trials to gain a picture or approximation of these properties. One can compare MCMC with sounding out unknown territory by taking simulated random walks. This modeling approach thus explores the behavior of a (complex) mathematical object, like a posterior distribution, with the help of the MCMC machinery. In a way, MCMC explores mathematical properties with the help of probabilistic and iterative means. One can see a *frequentist* element here sneaking in.

However, I want to make an additional point. The speed of MCMC is also an invitation to engage in an exploratory mode of modeling in the following sense. Modelers can work with incompletely specified models that contain parameters that get adjusted only in a feedback loop where model behavior is observed and modified. Researchers do not need to determine parameters from the beginning; rather, they can adapt them during the process to obtain a better match. For Bayesian modeling, MCMC made exploration on this level feasible. With the help of adjustable parameters, a model can be specified in flexible ways. The MCMC trick brings this flexibility to Bayesian modeling.

A short remark on the timeline. Typically, computational modeling of this explorative sort will be done when computational capacity is easily and cheaply accessible—including software packages (see section 5 below). On expensive mainframe machines, researchers tend to run only their best models with their best guesses. This accessibility condition started to be fulfilled in many labs and offices from the early 1990s onward; and this coincides with the timeline displayed in figure 1.

However, the exploratory–iterative mode affects the Bayesian rationale. The core of Bayesian epistemology, indeed the defining feature for many philosophers, is the subjective stance. The modeling process starts out with one's degrees of belief. We have seen, however, that

[25] I owe this formulation to an anonymous reviewer who helpfully inquired about senses of exploration.

this characteristic of Bayesian epistemology is fading away over the course of the development of MCMC approaches. Priors now appear as part of the adaptation machinery.[26] Importantly, these parameters are loose their interpretation as prior knowledge. To the extent that they are treated like adjustable parameters, the resulting values no longer express (degrees of) *prior* belief, but rather correspond to an overall fit of model and data, *resulting* from the exploratory–iterative process of modeling. In a nutshell, *the priors cease to be prior*.

I have presented the argument over how using MCMC as a tool undermines the perceived rationality of Bayesianism. I conclude this section by backing up this argument with a second line of thought that supports the claim about the exploratory–iterative nature of MCMC. Now is the time to recall the earlier promise, and address the second condition for MCMC. The one that looks innocent but is not: namely, that the Markov chain has reached equilibrium. The results MCMC provides take for granted that the Markov chain has reached stationarity before sampling. If the chain runs one million steps, is that enough? Or, if it is not quite in equilibrium, how does that play out in terms of error bars? Answering these questions is arguably the most important and intricate problem in the validation of MCMC results.

First of all, there are various approaches that try to implement a computational forward strategy: simulate the chain and observe whether it has reached stationarity. This sort of observation remains shaky because there might be relevant areas that the chain has not yet visited, or not visited with significant frequency. Maybe waiting twice as long will change the observed distribution significantly. As has been mentioned above, the effectiveness of MCMC relies essentially on how quickly Markov chains converge to their stationary distribution, sometimes called fast mixing. The speed of mixing is relative to the complexity of the space the random walk has to explore. The important question is: Exactly how quickly does the chain actually converge?

Answering this question is crucial for any assessment of MCMC results. Fast mixing and the rate of convergence have been identified as an important research topic being tackled by some of the most prominent researchers in stochastics and statistics.[27] Despite a growing number of

[26] Consider the chapter by Datta and Sweeting (2005) on matching priors that are used to adjust priors *so that* the posterior distribution has the desired properties. Instances such as this abound—they are unavoidable when working in an iterative–exploratory mode.

[27] The monographs by Levin et al. (2009) and Aldous and Fill (2002) document the settled state of the art; Diaconis (2013) provides an outlook on current progress and challenges.

results and insights, there is still a large lacuna regarding the behavior of chains that move in large continuous spaces, as is typical with Bayesian posteriors. In fact, this is the downside of MCMC-enhanced modeling flexibility: The menagerie of MCMC-enhanced models is growing, whereas knowledge about convergence speed is still lacking. There may be a chance for a mathematical theory to eventually provide such a footing; yet up to now, no strong results exist. Diaconis reasons that the market may be populated by many applied MCMC algorithms that perform well, and that their careful analysis might present useful hints for directing the mathematical research toward why these algorithms behave so well (2009, p. 195)—or toward why they do not. Diaconis has no illusions about how limited the range of mathematical accounts of the validation problem is. He is alarmed by the tendency to build excessively complex models for which, thanks to MCMC, the Bayesian machinery still works, whereas considerations (about MCMC) that could help regarding validation are largely missing.

This exemplifies a problem whose significance goes beyond the case of statistics and Bayesianism. Namely, a technology-based mode of mathematical modeling pushes the limits of modeling so that questions of validation can be addressed only by quasiempirical means—that is, by observing the performance of the models. This state of affairs is endemic in computational modeling. Many researchers resort to a kind of quasiempirical forward strategy—that is, they explore via simulations how the model will behave under varying initial conditions. Carlin and Louis, for an instance from Bayesian statistics, argue:

> The most basic tool for investigating model uncertainty is the sensitivity analysis. That is, we simply make reasonable modifications to the assumption in question, recompute the posterior quantities of interest, and see whether they have changed in a way that has practical impact on interpretations or decisions. (2000, p. 194)

This validation strategy—explore and observe variation in model behavior—can be found in many areas of computational modeling. It is characteristic of a field in which predictions are created in the iterative–exploratory mode. There is nothing wrong with these strategies; they just express that modeling happens under a condition of partial epistemic opacity where model behavior is not controlled by clear-cut assumptions, but rather by an assemblage of epistemic and instrumental components whose resulting behavior is adjusted.

5. Software

Software plays an important role in the upswing of Bayesian methods. A revolution from the perspective of professional mathematicians and statisticians might not necessarily have great impact on the methods practitioners use. As with any other instrument, the (perceived) quality of the instrument has to attract and hold the interest of an array of potential users; and, furthermore, must be usable given their level of expertise. Software packages have been—and still are—key for distributing the iterative–exploratory approach inherent in MCMC that is the computational backbone of Bayesian modeling.

Such software is not a neutral framework, because the options it offers and the algorithms it implements tend to steer in whatever directions statistical practices move (cf. Mira, 2005). Bayesian software has had two major effects: one the flip side of the other. Due to its usability, together with the easy accessibility of networked computers, it has triggered a stunning distribution of Bayesian modeling far beyond the ranks of those who had a Bayesian inclination before the 1990s turn. At the same time, many of these novices in statistical modeling are attracted by the software's capacity to deal with more complex models rather than by the standard rationale of Bayesianism.

When the great potential of MCMC began to become manifest, MCMC pioneers such as A. F. M. Smith realized that a software package was the missing ingredient that could turn Bayesian modeling into a widely used approach (Smith, 1988). This was exactly what David Spiegelhalter and his coworkers at the MRC Biostatistics Unit in Cambridge (UK) were developing. In 1991, they rolled out the BUGS program (short for Bayesian Statistics Using Gibbs Sampling). It was freely available and popularized Bayesian modeling tremendously. BUGS acted as a platform for Bayesian modeling by generating code for MCMC-based analyses of models that users could specify (see Gilks et al., 1994, Thomas et al., 1992). It featured uncertainty propagation in graphical structures; but the main point, of course, is that modelers could use the software to compute a posteriori distributions of their models without having to master the mathematics of MCMC.

Not much later in 1996, now under Nicky Best who had changed from Cambridge to Imperial College, London, the descendant WinBUGS was published—a version running under Windows reflecting the growing demand from the side of users who had no connection to special computing facilities but worked on (relatively small) desktop computers.[28] WinBUGS acted as an efficient popularizer, enlarging the variety (and complexity) of possible models as much as the variety of users. Textbooks such as Ntzoufras (2009) guided readers into using WinBUGS, with the selling point being that this free software "could fit complicated models in a relatively easy manner, using standard MCMC methods" (p. xvii). The entire book is devoted to WinBUGS, but time runs quickly in software development. By the end of the 2000s, the BUGS program had been turned into the open-source code OpenBUGS that is very similar to WinBUGS but also runs on Linux, Apple, and other Unix-related operating systems.[29] Importantly for computational modelers, OpenBUGS can be run from R and from SAS—that is, from the most common platforms of statistical analysis, thus creating a software environment for statistical modelers.

There is a plethora of packages that come into play on different levels. Some scientists use MCMC methodology by interfacing their data with a (more or less) complete tool for analysis like the BUGS family offers.[30] The BUGS family is not the only type of software package. Others invest work in developing their own customized MCMC simulations using software packages such as Mathematica or MathLab more as a generic tool kit. One important feature of "complete tool" software such as BUGS is the way one can make use of it. It not only provides a graphical interface, but also comes with a book of examples. Hence, users do not have to learn how it works—that is, how to specify their model case. Instead, they can build directly on particular examples. As Carlin puts it: "you don't read the manual; instead, you find the example that most nearly matches your situation, copy it, and modify it" (Kass et al., 1998, p. 94; cf. also Carlin, 2004).

Now, however, the flip side comes into play. MCMC has unresolved issues with convergence as we have seen above. Standard software has no guardrails that would prevent users from ignoring this issue. Jeff Gill (2008), for instance, recapitulates the enormous success of software packages in solving the needs of modelers but also warns:

> Unfortunately, these solutions can be complex and the theoretical issues are often demanding. Coupling this with easy-to-use software, such as WinBUGS and MCMCpack,

---

[28] Lunn et al. 2000 is the standard reference for WinBUGS.

[29] See Lunn et al. (2009) and, for very brief historical remarks, the official website openbugs.net.

[30] Other popular software includes JAGS (Just Another Gibbs Sampler), Stan (developed at Columbia University), MCMCpack, bayesm, or the SAS MCMC. Their differences are of no concern here.

> means that there are users who are unaware of the dangers inherent in MCMC work. (p. xx)

The OpenBUGS website issues a "Health Warning": The programs are reasonably easy to use and come with a wide range of examples. There is, however, a need for caution. Knowledge of Bayesian statistics is assumed, including recognition of the potential importance of prior distributions; and MCMC is inherently less robust than analytic statistical methods. The fact that there is a (largely) unknown level of uncertainty should sound an alarm. However, there is no built-in protection against ignoring this fact.

Not surprisingly, the convergence problem is a matter addressed in some of the available packages. Interestingly, because a mathematical solution of the validation problem is out of reach, the software resorts to heuristic strategies to explore model behavior. The software AWTY (Nylander et al., 2008) provides a case in point. The acronym expands into "are we there yet?"—that is, has the chain reached equilibrium? The program is made for graphical exploration of convergence in the special case of Bayesian phylogenetics.

One can ask to what extent the lack of built-in guardrails poses an actual problem in statistical practice. This is hard to judge. It is not at all unlikely that practical methods are valid, although they cannot be fully justified mathematically. In lieu of a reasoned judgment, I can only offer an impression. It looks like the standards of what counts as sound methodology are beginning to change. They are moving away from a mathematical paradigm tied to proof to a computational paradigm where skillful modification is the key.

The wide uptake of Bayesian methods reflects the social organization of the field. The number of users was able to grow dramatically because—thanks to the software—these users do not need to be experts in neither statistics, nor the mathematics of MCMC. Bayesian methods have become a pragmatic and flexible tool in statistical practice. At the same time, this flexibility leads to an erosion of the original rationale—whether frequentist subparts are utilized, or whether priors express a meaningful subjective stance is not per se important if the machinery works.

6. Prediction and Pragmatism

Bayesian approaches are a success story in statistics that began in the 1990s. I have argued that this story pivots on the codevelopment of computational methods and a class of models that, when working together, made predictions possible. From a methodological perspective, MCMC was the key factor; from a social perspective, the widely available software that runs on networked computers was a key contribution to the success in practice. In a nutshell, Bayesian statistics evolved into an exploratory–iterative culture of prediction. This evolution affected Bayesian rationality in an important way. From a philosophically somewhat clean approach to a pragmatic tool that comes with more philosophical laissez-faire. This pragmatic turn has the potential to fundamentally affect the philosophy of statistics. The principled interpretation of Bayes' rule has been the major bone of contention. Statisticians are aware of this transformation, not least because elements that formerly counted as incompatible now come together in predictive practices. How the new situation should be captured conceptually is not yet clear. Some scholars want to restrict the title "Bayesian" to approaches that stick to the Bayesian principle. They are critical of the newer, prediction-oriented approaches. Others, who still perceive themselves as Bayesians, side with prediction making.[31] Here is a sample of responses from statisticians.

According to Bradley Efron,[32] classical frequentist and Bayesian approaches work together and mutually *complement* each other in computer modeling. Especially when analyzing large amounts of ("big") data—according to Efron (2005)—it is often hopeless to construe priors in a subjective way. Sander Greenland (2010) argues that Efron's stance on the mutually complementing virtues is not correct and that it would be better to use the term "ecumenism" to describe how statistical methods come together.[33] He traces this back to G. E. P. Box's (1983) plea for ecumenism. Despite its prominent advocates—according to Greenland—ecumenism has

[31] To do justice to the full richness of Bayesian approaches, it would be necessary, as noted above, to include approaches such as "objective" and "evidentialist" Bayes. Instead of entering a more detailed appraisal (see footnote 2), this paper prefers a simplistic approach to Bayesianism and fully concentrates on the computational perspective.

[32] Efron is famous for the bootstrap method that works in a frequentist guise. Therefore, he is presumably a significant witness in favor of Bayesian approaches. Moreover, he has repeatedly made claims that statistical inference has been transformed through computer use (explained in book length in Efron and Hastie, 2016).

[33] Gill (2008) also favors "ecumenism."

not yet had a large impact on the teaching or practice of statistics.[34] Robert Kass is another prominent statistician who reflects on the ongoing changes in a conceptual way. He advocates what he calls "statistical pragmatism," a position that sees modeling as the core activity (Kass 2011). He makes a careful attempt to sketch the common ground between Bayesian and frequentist positions regarding how statistical models are connected with data. Thus, the dynamics of computational modeling seem to be a uniting feature of formerly separated camps of philosophy of statistics: "The loyalists of the 1960s and 1970s failed to realize that Bayes would ultimately be accepted, not because of its superior logic, but because probability models are so marvelously adept at mimicking the variation in real-world data" (Kass, cited according to McGrayne, 2011, p. 234).[35] Steven Goodman (2011) disagrees, because Kass' pragmatism looks like a mere truce rather than a new foundation. Also commenting on Kass, Hal Stern (2011, p. 17) worries "more broadly that pragmatism might appear to reinforce the notion of statistics as a set of techniques that we 'pull off the shelf' when confronted with a data set of a particular type." Finally, Andrew Gelman (2011, p. 10) observes that this pragmatism, though thriving on the flexibility of methods to obtain calibration between model and data, is still objective. In sum, notions such as complement, truce, ecumenism, or pragmatism show that statisticians grapple with reflecting on what happens in practice and whether this makes a discussion about the foundations dispensable or, on the contrary, downright demands it.

References

Adams, Douglas (1979). *The Hitchhikers' Guide to the Galaxy*. Pan Books.

Aldous, D. and Fill, J. (2002). *Reversible Markov chains and random walks on graphs*. Unpublished monograph. Accessible online: https://www.stat.berkeley.edu/~aldous/RWG/book.pdf (accessed Oct 20, 2020).

Bélisle, C. J. P., Romeijn, H. E., and Smith, R. L. (1993). Hit-and-Run Algorithms for Generating Multivariate Distributions. *Mathematics of Operations Research 18*, 255–266.

Box, G. (1983). An apology for ecumenism in statistics. In: *Scientific Inference, Data Analysis, and Robustness* (G. Box, T. Leonard, and C.-F. Wu, eds.) 51–84. Academic Press, New York.

Carlin, Bradley P. (2004). Whither Applied Bayesian Inference? in: Andrew Gelman and Xiao-Li Meng (eds.): *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. (honoring Donald Rubin)*, Chichester, England: Wiley pp. 279–284

[34] Greenland further acknowledges that this theme is not new, but also has been brought up repeatedly by Good (1983), Diaconis and Freedman (1986), which includes a discussion, or Samaniego and Reneau (1994).

[35] This capability is based heavily on adaptable parameters, especially on priors that can be changed to increase the ability of a model to mimic the data—quite in line with our prior analysis of MCMC.

Carlin, Bradley P. and Louis, Thomas A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall,.

Chen, M.-H., and Schmeiser, B.W. (1996). General Hit-and-Run Monte Carlo Sampling for Evaluating multidimensional integrals. *Operations Research Letters 19*, 161–169.

Clifford, P. (1993). Discussion Statement, in: Discussion on the meeting on the Gibbs sampler and other Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society B*, *55*(1), 53–102.

Corfield, David and Williamson, Jon (eds. (2001). *Foundations of Bayesianism*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Daston, Lorraine (1988). *Classical Probability in the Enlightenment*, Princeton University Press.

Datta, Gauri Sankar and Sweeting, Trevor J. (2005). Probability Matching Priors, in: Dey, D. K., and C. R. Rao (eds.): *Bayesian Thinking: Modeling and Computation. Handbook of Statistics 25*, Amsterdam: Elsevier, pp. 91–114.

Diaconis, Persi (2013). Some Things We've Learned (about Markov chain Monte Carlo), *Bernoulli* 19(4), 1294–1305.

Diaconis, Persi (2009). The Markov Chain Monte Carlo Revolution. *Bulletin of the American Mathematical Society, 46(2), pp. 179-205.*

Diaconis, Persi (1998): A Place for Philosophy? The Rise of Modeling in Statistical Science. *Quarterly of Applied Mathematics, LVI*(4), 797–805.

Diaconis, P. and Freedman, D. (1986). On the Consistency of Bayes Estimates (with discussion). *Annals of Statistics*, 14, pp. 1–67.

Earman, J. (1992). *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA and London, England: The MIT Press.

Efron, Bradley (2005). Bayesians, Frequentists, and Scientists. *Journal of the American Statistical Association 100*(469), presidential address, 1–5.

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Gelman, A. (2011). Bayesian Statistical Pragmatism. *Statistical Science*, 26(1), 10–11.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions of Pattern Analysis and Machine Intelligence,* 6, 721–741.

Gilks, Wally R., Thomas, Andrew and Spiegelhalter, David J. (1994). A Language and Program for Complex Bayesian Modelling, *The Statistician*, 43(1), pp. 169-177.

Gill, Jeff (2008). *Bayesian Methods, A Social and Behavioral Sciences Approach*, Boca Raton, FL: Chapman and Hall.

Good, I. J. (1983). *Good Thinking*. Univ. Minnesota Press, Minneapolis.

Goodman, S. N. (2011). Discussion of "Statistical Inference: The Big Picture" by R. E. Kass, *Statistical Science*, 26(1), 12–14.

Greenland, Sander (2010). Comment: The Need for Syncretism in Applied Statistics. *Statistical Science*, 25(2), 158–161.

Hacking, Ian (1990). The *Taming of Chance*. Cambridge, UK: Cambridge University Press.

Hacking, Ian (2001). *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.

Häggström, Olle (2002). *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press.

Hastings, W.K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, *57*(1): 97–109.

Howson, Colin, and Urbach, Peter (1993): *Scientific Reasoning: The Bayesian Approach*, Chicago and La Salle, IL: Open Court (2nd ed.).

Hughes, R.I.G. (1999). The Ising Model, Computer Simulation, and Universal Physics. In: M. Morgan and M. Morrison (eds.), *Models as Mediators*, pp. 97-145, Cambridge: Cambridge University Press.

Humphreys, P. (2004). *Extending Ourselves. Computational Science, Empiricism, and Scientific Method.* New York: Oxford University Press.

Kass, Robert E. (2011). Statistical Inference: The Big Picture, *Statistical Science*, 26(1), pp. 1–9.

Kass, Robert E., Carlin, Bradley B., Gelman, Andrew, and Neal, Radford M. (1998). Markov Chain Monte Carlo in Practice: A Roundtable Discussion, *The American Statistician*, *52*(2), 93–100.

Lenhard, Johannes (2006). Models and Statistical Inference: The Controversy Between Fisher and Neyman-Pearson, *British Journal for the Philosophy of Science*, 57, pp. 69–91.

Levin, D.A., Peres, Y. and Wilmer, E.L. (2009). Markov Chains and Mixing Times. Providence, RI: American Mathematical Society.

Lindley, D.V. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, Cambridge University Press.

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

Lunn, D., Spiegelhalter, D., Thomas, A. and Best, N. (2009). The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine, 28*: 3049–3082.

Mayo, Deborah G. and Spanos, Aris (2009). *Error and Inference. Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science*. Cambridge: Cambridge University Press.

McGrayne, Sharon Bertsch (2011). *The Theory That Would Not Die. How Bayes' rule cracked the enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy.* New Haven and London: Yale University Press.

Mira, Antonietta (2005). MCMC Methods to Estimate Bayesian Parametric Models, in: D.K. Dey and C.R. Rao (eds.), Bayesian Thinking: Modeling and Computation. Handbook of Statistics 25, Amsterdam: Elsevier, pp. 415-436.

Neal, Radford M. (1998). Philosophy of Bayesian Inference. URL (visited 1/2021): https://www.cs.toronto.edu/~radford/res-bayes-ex.html

Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: John Wiley and Sons.

Nylander, J.A., Wilgenbusch, J.C., Warren, D. L., and Swofford, D.L. (2008). AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. *Bioinformatics*, *24*(4), 581–583.

Pearl, Judea (1995). Causal Diagrams for Empirical Research *Biometrika, 82*(4), 669–710.

Porter, Theodore N. (1995). *Trust in Numbers. The Pursuit of Objectivity in Science and Public Life.* Princeton University Press.

Romeijn, Jan-Willem (2017). Philosophy of Statistics, *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/statistics/>.

Samaniego, F. J. and Reneau, D. (1994). Toward a Reconciliation of the Bayesian and Frequentist Approaches to Point Estimation, *Journal of the American Statistical Association*, 89, pp. 947–957.

Smith, A. F. M. (1984). Present Position and Potential Developments: Some Personal Views: Bayesian Statistics, *Journal of the Royal Statistical Society. Series A*, pp. 245-259.

Smith, A. F. M. (1988). What Should Be Bayesian About Bayesian Software? in: J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. F. M. Smith (eds.): *Bayesian Statistics 3*. Oxford, England: Clarendon Press, pp. 429–435.

Smith, A. F. M. and G. O. Roberts (1993). Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B*, *55*(1), 3–23.

Stern, H. (2011). Discussion of “Statistical Inference: The Big Picture” by R. E. Kass, *Statistical Science*, *26*(1), 17–18.

Talbott, William (2016). Bayesian Epistemology, *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/epistemology-bayesian/>.

Thomas, Andrew, Spiegelhalter, David J., and Gilks, Wally R. (1992). BUGS: A Program to Perform Bayesian Inference using Gibbs Sampling, in: J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith (eds.): *Bayesian Statistics 4.* Oxford, England: Clarendon Press, pp.837–842.

Titterington, Michael D. (2004). Statistical Modeling and Computation, in: Andrew Gelman and Xiao-Li Meng (eds.): *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives. (honoring Donald Rubin),* Sussex, England: Wiley, pp. 189–194.

Williamson, Jon (2005). *Bayesian Nets and Causality*. New York: Oxford University Press.

Zellner, A. (1988). A Bayesian Era in: J. M. Bernardo, M. H. Degroot, D. V. Lindley, and A. F. M. Smith (eds.): *Bayesian Statistics 3*. Oxford, England: Clarendon Press, pp. 509–516.