

# Reflections on Bayesian inference and Markov chain Monte Carlo

Radu V. CRAIU<sup>1</sup> , Paul GUSTAFSON<sup>2\*</sup> , and Jeffrey S. ROSENTHAL<sup>1</sup> 

<sup>1</sup>Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

<sup>2</sup>Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

*Key words and phrases:* Bayesian inference; Markov chain Monte Carlo.

*MSC 2020:* Primary 62F15; secondary 62C10.

*Abstract:* Bayesian inference and Markov chain Monte Carlo methods are vigorous areas of statistical research. Here we reflect on some recent developments and future directions in these fields.

*Résumé:* L'inférence bayésienne et les méthodes de Monte-Carlo par chaîne de Markov sont des domaines dynamiques de la recherche statistique. Les auteurs de ce travail présentent leurs réflexions au sujet de certains développements récents et d'orientations futures dans ces domaines.

## 1. INTRODUCTION

On the occasion of the 50th anniversary of the Statistical Society of Canada, we offer some reflections on research in the highly twinned areas of Bayesian inference and Markov Chain Monte Carlo (MCMC) methods. These areas continue to attract robust participation from the Canadian research community. As labelled by initials, each of Sections 2 through 4 offers thoughts from one of us. First, JSR and RVC survey computational topics related to the use of MCMC algorithms for Bayesian computation. Then, PG describes some Bayesian applications, specifically in the context of pandemic-related research.

## 2. MCMC IN HIGH DIMENSIONS (JSR)

MCMC algorithms are very widely used to explore and sample from a complicated high-dimensional target probability distribution  $\pi$ . Although other applications exist, the most common use of MCMC is in cases where  $\pi$  is a posterior distribution

$$\pi(\theta|\mathbf{y}_0) = \frac{f(\mathbf{y}_0|\theta)p(\theta)}{\int f(\mathbf{y}_0|\theta)p(\theta)d\theta}, \quad (1)$$

which is defined in terms of a prior distribution  $p$ , and a sampling density  $f$  that is indexed by a parameter  $\theta \in \Theta \subset \mathbf{R}^d$  and evaluated at the observed data  $\mathbf{y}_0 \in \mathcal{X}$ .

\* Corresponding author: [gustaf@stat.ubc.ca](mailto:gustaf@stat.ubc.ca)

The denominator in Equation (1) is usually mathematically intractable and impedes the calculation of posterior quantities such as

$$I = \int_{\Theta} h(\theta)\pi(\theta|\mathbf{y}_0)d\theta, \quad (2)$$

where  $h$  is a  $\pi$ -integrable function of interest. For instance, if  $d = 1$ , then taking  $h(\theta) = \theta^r$  in Equation (2) yields the  $r$ th moment of  $\pi$ , and taking  $h(\theta) = \mathbf{1}_{(-\infty, t]}(\theta)$  yields the cumulative distribution function (cdf) of  $\pi$  at a point  $t$ .

The most basic version of MCMC is the Metropolis algorithm (Metropolis et al., 1953). From a given state  $\theta$ , the algorithm proceeds by first proposing to move to a new state  $\omega$ , and then either accepting that proposal (i.e., moving to  $\omega$ ), or rejecting that proposal (i.e., staying at  $\theta$ ). The acceptance probability is given by  $\min[1, \pi(\omega)/\pi(\theta)]$ . If the proposal densities are symmetric (i.e., have the same probability of proposing  $\omega$  from  $\theta$  as of proposing  $\theta$  from  $\omega$ ), this procedure ensures that the resulting Markov chain will be reversible with respect to  $\pi$ , and thus have  $\pi$  as its stationary density. It then follows from standard Markov chain theory that, under mild irreducibility conditions, the probabilities and sample averages of the MCMC algorithm will converge respectively to the corresponding stationary distribution and expected values, thus facilitating sampling and estimation. The Metropolis and related algorithms have been extremely influential in Bayesian computation and led to many thousands of research articles exploring their theory and application; see e.g., Brooks et al. (2011) and the many references therein.

An important question about MCMC algorithms is how quickly they will converge to their stationary distribution. This has been the subject of much theoretical (e.g., Meyn & Tweedie, 1994; Rosenthal, 1995a,b, 1996, 2002; Roberts & Tweedie, 1999; Jones & Hobert, 2001, 2004; Baxendale, 2005) and practical (e.g., Gelman & Rubin, 1992) investigation, mostly focused on the specific question of how many iterations are required, for a particular statistical model and data, to get within a specified distance (say, 0.01) of stationarity in terms of a standard metric on probability distributions (say, total variation distance).

Generally, MCMC algorithms tend to converge reasonably quickly on the modest sized problems that statisticians have traditionally studied. However, the modern “big data” era, with models involving many thousands of parameters and millions of data points, presents new “curse of dimensionality” challenges. Thus, algorithms that are perfectly satisfactory in the former context may still fail in the latter context. This issue was driven home to me when I realized to my surprise that one of my best quantitative convergence bounds, proving that a modest 140 iterations suffices to get within 0.01 of stationarity for a certain complicated Bayesian model with 18 observations (Rosenthal, 1996), would nevertheless grow exponentially quickly as the amount of data grew to infinity.

As a result, statisticians have recently become more interested in the computer science concept of computational complexity (e.g., Cook, 1971), i.e., how the convergence times increase as the problem size (e.g., the amount of data) grows to infinity. Although various MCMC researchers had considered computational complexity to some extent in earlier MCMC analyses (e.g., Rosenthal, 1995b; Woodard, Schmidler & Huber, 2009a,b), this concept has taken on new prominence in recent years (e.g., Woodard & Rosenthal, 2013; Rajaratnam & Sparks, 2015; Yang, Wainwright & Jordan, 2016; Qin & Hobert, 2017; Yang & Rosenthal, 2017; Tawn, Roberts & Rosenthal, 2020). These new results are very promising, but they tend to apply only to specific models, or to require unverifiable assumptions, which limits their utility as a general solution to the curse of dimensionality.

In some cases, the problem of MCMC’s computational complexity can be avoided entirely by using, instead, an approximation method such as variational Bayes (Blei, Kucukelbir &

McAuliffe, 2017) or INLA (integrated nested Laplace approximation) (Rue et al., 2017). However, it is usually far preferable to find general methods of obtaining verifiably accurate and reliable MCMC samples even in high dimension. Indeed, I would argue that this is the central challenge of Bayesian computation today.

One approach in this direction is to use the results of Roberts and coauthors on the optimal scaling of proposal distributions (e.g., Roberts, Gelman & Gilks, 1997; Roberts & Rosenthal, 1998). These results involve speeding up the original Metropolis algorithm by a fixed power of the dimension  $d$  (e.g.,  $d$  or  $d^{1/3}$ ), and then proving that as  $d \rightarrow \infty$ , the resulting Markov process converges to a diffusion limit whose speed can then be optimized. The limiting diffusion no longer depends on dimension, which seems to imply that the computational complexity of the algorithm is equal to the order by which the chain was sped up. This intuition was formalized in Roberts & Rosenthal (2016), where it was proven that traditional random-walk Metropolis algorithms converge to stationarity in  $O(d)$  iterations, and Metropolis-adjusted Langevin algorithms converge in  $O(d^{1/3})$  iterations. However, these results were established only under the very strong assumptions required for the original diffusion limits, including that the target distribution factors into i.i.d. components (although in simulations they do appear to approximately hold much more generally (Roberts & Rosenthal, 2001)). These theoretical assumptions would never hold in practice, which leads to questions about how generally the corresponding results will hold in real MCMC applications.

So where does this leave us? We can say without hesitation that there has been a tremendous amount of success in using and verifying MCMC on moderately sized problems. However, when MCMC is used on much larger problems, it tends to suffer from either too slow convergence, or lack of reliability, or incomplete theoretical justification, or inaccurate approximation, or unrealistic assumptions. Thus, the overall goal of obtaining fast, accurate, reliable, verifiable MCMC sampling algorithms in general, very high-dimensional problems is not yet completely resolved. On the positive side, this vexing challenge provides many new research directions, which will surely occupy the MCMC and Bayesian computation communities for many years to come.

### 3. CHALLENGES IN BAYESIAN COMPUTATION (RVC)

#### 3.1. The Workhorse of Bayesian Computation

Although it continues to be referred to as a last-resort computational method (e.g., Thompson, 2011) to be used only when other numerical methods are ineffective, MCMC sampling has been widely adopted because of its ease of implementation and available software. Expanding on the discourse of Section 2, consider Hastings' generalization of the Metropolis sampler (Hastings, 1970), the so-called Metropolis–Hastings (henceforth, MH) algorithm, which is probably the most widely used MCMC sampler.

Assume that the state of the chain at time  $t$  is  $\theta_t$ . Given a user-defined proposal distribution  $q(\cdot|\theta_t)$ , the update rule to construct  $\theta_{t+1}$  is defined by the following two steps:

*Step 1* Draw a proposal  $\theta^*$  from the proposal density  $q(\theta|\theta_t)$ .

*Step 2* Set

$$\theta_{t+1} = \begin{cases} \theta^* & \text{with probability } \alpha, \\ \theta_t & \text{with probability } 1 - \alpha, \end{cases}$$

where

$$\alpha = \min \left\{ 1, \frac{\pi(\theta^*|\mathbf{y}_0)q(\theta_t|\theta^*)}{\pi(\theta_t|\mathbf{y}_0)q(\theta^*|\theta_t)} \right\}. \quad (3)$$

The pesky denominator in Equation (1) does not impede the calculation in Equation (3). However, the latter requires the ability to calculate the sampling density  $f(\mathbf{y}_0|\theta)$  for any parameter value  $\theta$ . The modern challenges posed by Bayesian computation have their roots in this apparently simple requirement.

### 3.2. Challenges Posed by Big Data

The increase in data volume presents an important initial challenge to classical MCMC procedures. For a sample of size  $N$  in the hundreds of thousands or millions, running an MCMC sampler for thousands of iterations becomes inefficient at best and impossible at worst since most samplers use at least  $O(N)$  operations to update the underlying Markov chain. Occasionally, the data volume is too large to be stored on a personal computer, making it impossible to update the Markov chain using Equation (3).

In other areas of statistical computation, algorithm designers have taken advantage of parallelization strategies in which a task is divided between a number of parallel “workers,” where a worker can be a processing unit, a computer core, etc. (e.g., Schervish, 1988; Schubert & Gertz, 2018). Alas, MCMC samplers are notoriously resistant to parallelization and require the design and adoption of new ideas and different approximation techniques. For instance, Suchard et al. (2010) propose using a GPU’s multiple processors to speed up a block Gibbs algorithm. Their approach exploits the GPU’s multiple cores, which allows computing tasks within each MCMC iteration to be parallelized. The proposed approach relies on the intrinsic synchronicity of the parallel tasks, i.e., all cores must complete their tasks before the chain can finish updating. This approach is difficult to generalize to other MCMC samplers where an update cannot be easily split among independent workers. Others have discussed parallel MCMC methods (Rosenthal, 2000; Laskey & Myers, 2003; Wilkinson, 2006), where each worker runs on the full dataset. However, these methods do not resolve memory overload and may require intensive communication between workers during the simulation, which can push the computation budget beyond realistic bounds. It has been recognized that synchronous coordination (i.e., workers must wait for each other to finish before moving on to the next step) or frequent communication between workers slows down computation significantly and should be avoided. Some progress on asynchronous updates in Gibbs samplers is reported in Terenin, Simpson & Draper (2020), albeit under relatively stringent conditions.

One strategy to reduce communication between workers and eliminate the need to store all the data on a single server is found in the embarrassingly parallel design in which the data are partitioned into  $K$  equally sized subsets called shards, each of which is analyzed independently by a different worker. Each worker uses an MCMC algorithm to draw samples from the posterior corresponding to that data shard, which is called a subposterior. Some essential MCMC-related questions regard (1) which subposterior distributions should be built for each shard and (2) how MCMC samples obtained from each subposterior should be combined to recover all information that would have been produced by an MCMC sample from the full posterior distribution. The subposteriors developed in the literature for the  $s$ th shard have the form

$$\pi^{(s)}(\theta|\mathbf{y}_0^{(s)}) \propto p(\theta)^{a_s} f(\mathbf{y}_0^{(s)}|\theta)^{b_s},$$

where  $\mathbf{y}_0^{(s)}$  is the  $s$ th data shard and  $a_s, b_s \in \mathbf{R}$  are user-defined. For instance, Wang & Dunson (2013), Neiswanger et al. (2013), Scott et al. (2016), and Nemeth & Sherlock (2018) use  $a_s = 1/K$  and  $b_s = 1$  for all  $s = 1, \dots, K$ , but differ in their approach to combining subsamples. Specifically, Neiswanger et al. (2013) approximate each subposterior using a kernel density estimator, Wang & Dunson (2013) use the Weierstrass transformation, Scott et al. (2016) use a weighted average with weights depending on subposterior variances, and

Nemeth & Sherlock (2018) use Gaussian processes. Another option is to use  $a_s = 1$  and  $b_s = K$  for all  $s = 1, \dots, K$  as in the likelihood-inflating algorithm (LISA) developed in Entezari, Craiu & Rosenthal (2018) for Bayesian additive regression trees. More recently, Changye & Robert (2019) have proposed  $a_s = \lambda_s/K$  and  $b_s = \lambda_s$  and used random forests and reweighted importance sampling to approximate subposteriors.

These divide-and-conquer methods are well justified for Gaussian (or mixtures thereof) posteriors and subposteriors, but theory is lacking otherwise. The most general “fix” is to use importance sampling so that each subposterior sample is properly weighted, but this strategy adds significantly to the computational burden. Extending divide-and-conquer techniques to models for non-i.i.d. data is also challenging because batches are not independent or even exchangeable.

Because the size of the data poses serious challenges, it is reasonable to consider using only a subset of the sample to speed up MCMC computation. One essential aim is to not alter the statistical properties of the sample when trimming it. In its simplest and most intuitive form, sample reduction can be easily understood when the model admits a sufficient statistic and the posterior that conditions only on it, instead of the full sample, provides the same information about the parameter. Alas, such simplicity is almost never available in complex models, so new ideas are needed.

Consider the log likelihood obtained by adding  $N$  independent terms, each corresponding to an independent item in the sample:

$$l_N(\theta|\mathbf{y}_0) = \sum_{i=1}^N l_i(\theta),$$

where  $l_i(\theta) = \log f(y_{0,i}|\theta)$ . It is tempting to take a random subsample of size  $r$  of the data items  $\{i_1, \dots, i_r\}$  and use  $l_r(\theta) = \sum_{j=1}^r l_{i_j}(\theta)$ . However, replacing  $l_N$  by  $l_r$  in the transition kernel of the Markov chain used to run, say, an MH sampling algorithm, alters the target distribution of the chain and can lead to large inferential errors. This concern can be alleviated, if an unbiased estimator of the likelihood is available, given the pseudomarginal approach developed in Andrieu & Roberts (2009). The authors show that the target distribution of an MH sampling algorithm remains unchanged if the likelihood is replaced by an unbiased estimator when computing the acceptance ratio in Equation (3). An important addendum is that the sampling efficiency of the new algorithm degrades as the variance of the ratio of the likelihoods increases (Andrieu & Vihola, 2015), thus discouraging a choice of  $r$  much smaller than  $N$ .

Some of these challenges are tackled in Quiroz et al. (2018), which specifies a subsampling-based approach in which a different subsample of size  $r$  is used in each iteration of the MCMC sampler. The authors introduce an estimator of the likelihood that is approximately bias-corrected and, to keep in check the variance of the likelihood ratio, control variates. Because the estimator is not exactly unbiased, the transition kernel of the MH chain deviates from the original one and has a perturbed target distribution. However, Quiroz et al. (2018) bound, for important general classes of models, the total variation distance between the perturbed chain’s target and the posterior distribution of interest,  $\pi(\theta|\mathbf{y}_0) \propto f(\mathbf{y}_0|\theta)p(\theta)$ , and make recommendations about choosing the size of the subsample.

Keeping with the “data trimming” theme, Huggins, Campbell & Broderick (2016) considered a different approach based on the concept of a Bayesian coresets. Here, a single subset of the data is selected and treated as “the data” for inferential purposes. The working assumption is that most large data are redundant, so it is possible to reduce their size while preserving their statistical properties. Bounds for the distance between the full data likelihood and the one corresponding to the coresets are derived theoretically. For Gaussian models, Huggins, Campbell & Broderick (2016) also derived bounds for the discrepancy between the full and reduced data posteriors, but similar results are difficult to quantify theoretically in more general cases.

### 3.3. Challenges Posed by Intractable Likelihoods

The modern statistician must deal with workflows of increasing complexity. Large studies not only lead to large volumes of data but can also provide answers to more complex questions. The latter require complex models that elude closed forms. Computer simulators of complex phenomena, e.g., hurricane-driven surges (Plumlee et al., 2021), groundwater modelling (Cui et al., 2018) or climate-change scenarios (Oyebamiji et al., 2015), exemplify generative models in which data can be generated for every configuration of model parameters, but the corresponding likelihood is not available.

When, for any parameter value  $\theta \in \mathbf{R}^q$ , synthetic data  $\mathbf{y} \sim f(\mathbf{y}|\theta)$  can be generated from the model, one can still conduct a Bayesian analysis. We discuss here two computational approaches that have gained considerable momentum in recent years: approximate Bayesian computation (ABC) (Marin et al., 2012; Baragatti & Pudlo, 2014; Drovandi, 2018; Sisson, Fan & Beaumont, 2018a) and Bayesian synthetic likelihood (BSL) (Wood, 2010; Drovandi et al., 2018; Price et al., 2018). Both algorithms are effective when combined with MCMC sampling schemes to produce samples from an approximation of the posterior.

In its simplest form, ABC is an accept/reject sampler. Given observed data  $\mathbf{y}_0$ , a user-defined threshold  $\epsilon > 0$ , a distance  $d : \mathbf{R}^p \times \mathbf{R}^p \rightarrow \mathbf{R}_+$ , and a summary statistic  $S(\mathbf{y}) \in \mathbf{R}^p$ , the algorithm has the following steps:

- S1 Sample  $\theta^* \sim p(\theta)$  and synthetic data  $\mathbf{y} \sim f(\mathbf{y}|\theta^*)$ .
- S2 If  $d(S(\mathbf{y}), S(\mathbf{y}_0)) \leq \epsilon$ , then accept  $\theta^*$  as a sample from the approximate posterior  $\pi_\epsilon(\theta|S(\mathbf{y}_0))$ , the marginal of the joint distribution

$$\pi_\epsilon(\theta, \mathbf{y}|S(\mathbf{y}_0)) \propto p(\theta)f(\mathbf{y}|\theta)\mathbf{1}_{\{d(S(\mathbf{y}), S(\mathbf{y}_0)) < \epsilon\}}. \tag{4}$$

When  $S$  is a sufficient statistic and  $\epsilon = 0$ , the approximate posterior is the true posterior, i.e.,  $\pi_\epsilon(\theta|S(\mathbf{y}_0)) = \pi(\theta|\mathbf{y}_0)$ . To verify, let us work under the simplifying assumption that both  $\theta$  and  $\mathbf{y}$  take discrete values. One can easily see that

$$\Pr(\theta = \theta_0|S(\mathbf{y}_0)) \propto p(\theta_0)\Pr(S(\mathbf{y}) = S(\mathbf{y}_0)|\theta = \theta_0) \propto \pi(\theta_0|\mathbf{y}_0)$$

which holds because  $S$  is sufficient and  $\epsilon = 0$ .

The remarkable feat of exploring an approximation of the posterior even when the likelihood is intractable has generated a lot of interest, as demonstrated by the large number of articles and ideas, all of which simply cannot be discussed here. Instead, I focus on a few essential developments. The practical implementation of the ABC algorithm requires choosing a number of simulation parameters, e.g.,  $d$ ,  $\epsilon$ , or  $S$ . Theory-backed recommendations for the choice of  $S$  can be found in Fearnhead & Prangle (2012) and Prangle (2015). More radically, Bernton et al. (2017) bypass the need to select the statistic  $S$  by computing the Wasserstein distance between the empirical distributions of the observed and synthetic data.

In the absence of information about model parameters, the prior and posterior distributions may have nonoverlapping regions with significant mass. Hence, parameter values that are drawn from the prior, as in S1, will rarely be retained. Recognizing this, Marjoram et al. (2003) proposed an ABC–MCMC algorithm that relies on building an MH transition kernel with the state space  $\{(\theta, \mathbf{y}) \in \mathbf{R}^q \times \mathcal{X}^n\}$ , proposal distribution  $q(\theta|\theta_t)f(\mathbf{y}|\theta)$  at iteration  $t$ , and the target  $\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0)$  given in Equation (4). The acceptance probability in Equation (3) can be computed exactly because the intractable terms involving the likelihood  $f(\mathbf{y}|\theta)$  cancel out. There are a few alternatives to Marjoram et al.’s sampler, motivated by low acceptance probabilities for small values of  $\epsilon$ . For instance, Lee, Andrieu & Doucet (2012) note that

$$\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0) \propto P(d(S(\mathbf{y}), S(\mathbf{y}_0)) < \epsilon|\theta) \tag{5}$$



and replace  $\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0)$  in the calculation of the acceptance probability with an unbiased estimator  $J^{-1} \sum_{j=1}^J \mathbf{1}_{\{d(S(\mathbf{y}_j), S(\mathbf{y}_0)) < \epsilon\}}$ , where  $J \geq 1$  and each  $\mathbf{y}_j$  is independently simulated from  $f(\mathbf{y}|\theta)$ . The acceptance probability tends to increase compared to Marjoram et al.'s sampler, but in order to keep the variance of the estimator under control, a large number of pseudo-data generations may be needed to produce stable estimates of the tail probability in Equation (5) when  $\epsilon$  is small. Other MCMC designs suitable for ABC can be found in Bornn et al. (2014). Sequential Monte Carlo (SMC) samplers have also been successfully used for ABC (Sisson, Fan & Tanaka, 2007; Lee, 2012; Filippi et al., 2013) and rely on a user-specified decreasing sequence  $\epsilon_0 > \dots > \epsilon_J$ . A comprehensive coverage of ABC-related theory and computational techniques can be found in Sisson, Fan & Beaumont (2018b) and the references therein.

An alternative approach to bypass the intractability of the sampling distribution is proposed in Wood (2010). This approach is based on the working assumption that the conditional distribution of a user-defined statistic  $S(\mathbf{y})$  given  $\theta$  is Gaussian with a mean  $\mu_\theta$  and a covariance matrix  $\Sigma_\theta$ . The synthetic likelihood (SL) procedure assigns to each  $\theta$  the likelihood  $\text{SL}(\theta|s_0) = \mathcal{N}(s_0; \mu_\theta, \Sigma_\theta)$ , where  $s_0 = S(\mathbf{y}_0)$  and  $\mathcal{N}(x; \mu, \Sigma)$  denotes the density of a normal random variable with a mean of  $\mu$  and a covariance of  $\Sigma$ . SL can be used for maximum likelihood estimation, as in Wood (2010), or within a Bayesian paradigm, as proposed in Drovandi et al. (2018) and Price et al. (2018). The latter work proposes sampling the approximate posterior generated by the BSL approach,  $\pi_{\text{SL}}(\theta|s_0) \propto p(\theta)\text{SL}(\theta|s_0)$ , using an MH sampler. Direct calculation of the acceptance probability is not possible because the conditional mean and covariance are unknown for any  $\theta$ . However, both can be estimated from  $m$  statistics  $s_1, \dots, s_m$  sampled from their conditional distribution given  $\theta$ . More precisely, after simulating  $\mathbf{y}_i \sim f(\mathbf{y}|\theta)$  and setting  $s_i = S(\mathbf{y}_i)$  for  $i = 1, \dots, m$ , one can compute

$$\hat{\mu}_\theta = \frac{\sum_{i=1}^m s_i}{m} \quad \text{and} \quad \hat{\Sigma}_\theta = \frac{\sum_{i=1}^m (s_i - \hat{\mu}_\theta)(s_i - \hat{\mu}_\theta)^\top}{m - 1}, \quad (6)$$

so the SL is

$$\text{SL}(\theta|s_0) = \mathcal{N}(s_0; \hat{\mu}_\theta, \hat{\Sigma}_\theta). \quad (7)$$

An MH algorithm designed to sample from the posterior  $\pi_{\text{SL}}(\theta|s_0)$  is updated through the following steps at iteration  $t$ ;

SL1 Sample  $\theta^* \sim q(\theta|\theta_t)$ .

SL2 Sample  $\mathbf{y}_1, \dots, \mathbf{y}_m \stackrel{\text{i.i.d.}}{\sim} f(\mathbf{y}|\theta^*)$  and compute  $\mu_{\theta^*}$  and  $\Sigma_{\theta^*}$  as in Equation (6) to obtain  $\text{SL}(\theta^*|s_0)$ .

SL3 Set  $\theta_{t+1} = \theta^*$  with probability  $\alpha = \min \left\{ 1, \frac{q(\theta_t|\theta^*)\pi_{\text{SL}}(\theta^*|s_0)}{q(\theta^*|\theta_t)\pi_{\text{SL}}(\theta_t|s_0)} \right\}$  and  $\theta_{t+1} = \theta_t$  with probability  $1 - \alpha$ .

Running MCMC samplers for either ABC or BSL involves generating multiple pseudosamples and can become extremely costly in situations where data are high-dimensional and very large or expensive to generate, like in the hurricane path or climate change examples. Anticipating that complex models are usually motivated by big data, Levi & Craiu (2022) proposed strategies to minimize the amount of pseudodata that needs to be simulated. When computing MH acceptance ratios, the authors recommend reusing some of the proposals  $(\theta, \mathbf{y})$  from the chain's history. This modification of the chain's kernel reduces computation time by orders of magnitude but perturbs the target distribution. Their theoretical developments demonstrate that this error can be controlled in independent Metropolis samplers.

### 3.4. Conclusion

The challenges posed by Bayesian computation are important and require classical methods to be significantly reframed. I have discussed a number of methods that address two types

of challenges: big data and intractable likelihoods. These two challenges cannot always be neatly separated and, in fact, I expect to increasingly see problems that combine the two. My discussion has not included important classes of approximation for Bayesian inference that eliminate the need for Monte Carlo sampling altogether, e.g., variational Bayes (Blei, Kucukelbir & McAuliffe, 2017) or INLA (Rue et al., 2017) (both also mentioned in Section 2). While these methods can be extremely efficient in certain models or when aimed at particular applications, they lack the level of generality exhibited by some of the other methods discussed so far.

Clearly, our efforts to expand the toolbox of Bayesian computation in the modern era are just revving up. Much remains to be explored, from new conceptual ideas to efficient and automatic implementations. I urge those willing to participate in this adventure to be more accepting of the idea that approximations are unavoidable when dealing with challenges like the ones I described here, but to also keep in mind that the errors incurred must be theoretically and practically controlled under any scenario of interest.

#### 4. BAYES IN THE TIME OF COVID (PG)

In contemplating the direction of my contribution to this article, I considered looking back to comment on the evolution of Bayesian analysis over recent decades. I also pondered looking forward to speculate on where this field might be headed. In the end, however, I decided to, instead, hone in on the present. While the global pandemic has upended society, Bayesian methods have been useful within the scientific response to the pandemic. I briefly describe several examples of this utility, drawing on the work I have been involved in as well as that of others.

##### 4.1. Uncertainty Arising in Diagnostic Testing

Most diagnostic tests used in medicine are imperfect. At least qualitatively, it is well known that false-positive and false-negative test results can occur. Quantitatively, epidemiologists and statisticians strive to react appropriately to this reality. In typical-use settings, a diagnostic test is evaluated methodically before widespread use and the manufacturer states the test's performance characteristics as part of a "package insert." This would be expressed as values for sensitivity and specificity, the chance that a correct test result ensues for a true positive and a true negative subject, respectively.

At the start of the pandemic, diagnostic tests for COVID-19 were developed and widely deployed with amazing haste. For instance, in the case of PCR testing of nasopharyngeal swabs to diagnose current infection, Canadian provinces started posting daily reports on their case-finding efforts around the beginning of February 2020. Soon after, by April 2020, the first serological tests for prior infection were being used in scientific studies (so-called "sero-surveys"). In both cases, and because of the necessary haste, testing data were collected and analyzed with less than the typical understanding of the diagnostic test's performance. Thus, statistical efforts were made to acknowledge the uncertainty about performance, a task for which Bayesian methods are well, if not singularly, suited.

Burstyn, Goldstein & Gustafson (2020) presented an early instance of an analysis acknowledging PCR test imperfection. (The first version of this article was posted to medrxiv.org on 11 April 2020.) The authors showed adjusted epidemic curves for both the province of Alberta and the city of Philadelphia for March 2020. More technically, they gave Bayesian point and interval estimates for the daily number of true positives among those receiving a test, as distinct from the daily number of reported positives.

In formulating the underlying Bayesian model, the authors made an important design choice to let the unknown test sensitivity vary by day, whereas the unknown test specificity was presumed static. To exemplify matching the prior specification to the scientific context at



hand, I elaborate briefly on this point, with pertinent discussions in Burstyn, Goldstein & Gustafson (2020) and Günther et al. (2021). It is well understood that PCR technology has effectively perfect “technical” sensitivity, i.e., virtually any small amount of live virus on the swab will light up the machine. The swabbing, however, is the weak link. Even a swab done by a highly trained health practitioner may not capture any virus particles if the (truly infected) test subject has low virus levels in the nasal cavity. Consequently, effective sensitivity will be lower among recently infected individuals compared to those with more established infections. Extrapolating further, when testing a largely asymptomatic population, true positives will tend to be recently infected, so sensitivity will be lower. However, if only those with respiratory symptoms are eligible for a test, then the true positives will tend to have more established infections, resulting in a higher sensitivity. It is then not reasonable to expect a globally valid package-insert value for test sensitivity, since the nature of the population being tested is critical. Moreover, particularly early in the pandemic, there were temporal changes in eligibility for a test (based on symptoms, say) due to kit availability, lab capacity, and other considerations. Hence, a prior specification allowing a smooth change in test sensitivity over time was an important ingredient in the analysis.

Turning now to testing for prior infection, in April 2020 a controversy arose in both scientific and lay circles regarding an early sero-survey conducted in Santa Clara, California. The initial preprint version of Bendavid et al. (2021) attracted much attention and criticism. A brief version of one concern is as follows. Package-insert values of test sensitivity and specificity,  $\tilde{s}_n$  and  $\tilde{s}_p$ , are inputs to an analysis inferring lifetime prevalence of infection from a sample. With a truly rare infection, even if  $\tilde{s}_p$  is close to 1, the estimated prevalence can vary strongly under a small perturbation to  $\tilde{s}_p$ . Because of the urgent need to roll out tests and studies at warp speed, validation studies to determine diagnostic test properties were based on relatively few (known negative and known positive) test specimens, and a problem ensued.

Gelman & Carpenter (2020) reviewed this challenging situation and provided a careful Bayesian analysis acknowledging the uncertainty in test specificity (and sensitivity). These authors demonstrated much greater *a posteriori* uncertainty about the lifetime prevalence in the study population relative to the initial version of Bendavid et al. (2021). Gelman and Carpenter also illustrated other modelling features that can be brought to bear quite simply in a Bayesian framework. Of note, they showed how a hierarchical Bayesian analysis can incorporate multiple validation studies, each of which involves similar but not identical test characteristics.

#### 4.2. Uncertainty Arising from Unknown Testing Patterns

Early in the pandemic, it was particularly challenging to learn infection rates and consequently infection fatality rates. Even in jurisdictions able to enumerate deaths due to COVID-19 infection relatively well, there were probably many undiagnosed infections. In terms of the infection fatality rate then, the numerator could be estimated reliably. However, particularly before the availability of serological tests for past infection, reasonably estimating the denominator was challenging.

Wu et al. (2020) presented a Bayesian analysis targeting the total number of infections in the United States up to mid-April 2020. As part of the authors’ modelling of state-level testing data, they built a defensible prior distribution describing relationships between the severity of respiratory symptoms and the likelihood of being tested. In common with Gelman & Carpenter (2020), they also used a prior distribution to adjust for the imperfect diagnostic test.

The principal finding of Wu et al. (2020) was that the total number of cases was likely between 3 and 20 times greater than the number diagnosed. Of course, this is a very wide credible interval. However, this width is appropriate given the information available at the time. In common with Gelman & Carpenter (2020), a thoughtful Bayesian analysis of the available

information did not yield a sharp answer. This is a feature though, not a bug. The principled propagation of uncertainty afforded by Bayes' theorem implies that sharper inference would not be justified.

In a related vein, Campbell et al. (2021) developed a Bayesian model admitting both surveillance data (tests for current infection within the confines and recommendations of public health surveillance at the time and place in question) and sero-survey data (tests for previous infection in a random sample conducted in the target time and place). The surveillance data were handled by modelling the association between infection status and testing status via an unknown parameter describing the extent to which the testing is preferential (i.e., honing in on those infected) rather than random. The authors ascribed a hierarchical prior distribution to these jurisdiction-specific preferential testing parameters. By applying this methodology to an evidence synthesis for European countries in Spring 2020, the infection fatality rate was estimated to be 0.53% (the posterior median) with an accompanying 95% credible interval of (0.39%, 0.69%). This inference is quite compatible with other estimates targeting European settings at about the same point in the pandemic but using different data sources and methods.

### 4.3. Just Turn the Bayesian Crank, or Understand it as Well?

In the examples alluded to above, Bayesian inference can be “plug and play.” Once the heavy lifting is done in making defensible model and prior assertions, these can be encoded in a Bayesian software package (see Section 5). Then, data go in and posterior inference comes out. Scientifically, this can be the end of the narrative arc. One has a general assurance that reported estimates and posterior uncertainties are founded in principle based on the combined information content of the data, the supplied model, and the prior assertions.

As a specific example of ending the narrative, Burstyn, Goldstein & Gustafson (2020) presented the posterior distribution of the (time-varying) sensitivity and (static) specificity of the diagnostic test. Focusing on the Alberta data, the posterior distribution of test sensitivity was very similar to the prior distribution, i.e., the data did not provide additional information about the false-negative rate. On the other hand, and perhaps curiously, the posterior distribution of specificity was far more concentrated than the prior distribution. The former concentrated above 99.5% specificity with a mode of 100%, despite the latter being uniformly spread between 95% and 100%. The data were quite certain that the specificity was very close to perfect. An applied user can simply stop here and take this away as useful (or not so useful, in the case of the finding for sensitivity) knowledge. However, the mathematical scientist will naturally ask: why are the findings such?

It turns out that the explanation for the above findings arises from the partial identification that underlies this Bayesian model. The concept of partial identification is long studied, particularly through a frequentist lens and with an econometrics slant. Manski (2003) is a well-known reference. More recently, I have attempted to give a somewhat thorough Bayesian treatment of the topic (Gustafson, 2015). In essence, in lower information settings such as those described above, not all the parameters are uniquely determined by the law of the observable data. However, those not uniquely determined may be subject to inequalities in terms of those that are. By elucidating these inequalities, one can understand directly how the Alberta data said so little about test sensitivity and so much about specificity.

As it happens, such an understanding of what lurks behind the crank is also considered in Campbell et al. (2021). Tucked away in a supplement is a mathematical elucidation of the partial identification structure applicable to the surveillance data part of the model. (The sero-survey component is much simpler, with those tested presumed to be a random sample from the population.) Again, by elucidating inequalities in the parameter space, one sees that, depending on the configuration of infection rates and the extent of preferential sampling across jurisdictions,

the surveillance data contribute differently to inference about the infection fatality rate. We can go beyond simply taking the posterior distribution as a *fait accompli*.

#### 4.4. Looking Ahead

Even before the onset of the pandemic, the role of statistical modelling and inference in improving the human condition was, arguably, on the upswing. As we all fervently wish for the pandemic to appear in the rear-view mirror, it seems likely that recent experiences will expedite this trend further. Because in large part of computational advances, the Bayesian toolbox proved itself to be at the ready as part of the pandemic response. This success should spur efforts to refine the toolbox further and to apply it more widely in the years to come.

### 5. DISCUSSION

To conclude this article, we briefly comment on the relationship between algorithms (the focus of Sections 2 and 3) and applications (the focus of Section 4). Clearly, the former enable the latter. That said, it is interesting to note how this enabling works. There are long-standing efforts to build software that can permit users to compute posterior quantities, without relying on users' expert knowledge of algorithms. While perhaps we are not fully there yet, the ideal is that an applied user will just declare a model and a prior, input the data, and press the "compute posterior" button.

By no means do we attempt to mention all of the software for Bayesian inference here. We note, however, that the BUGS (Bayesian inference using Gibbs sampling) software package (Gilks, Thomas & Spiegelhalter, 1994; Lunn et al., 2000; Lunn et al., 2009) and its continuation (roughly speaking), the JAGS (just another Gibbs sampler) software package (<https://mcmc-jags.sourceforge.io>), represent three decades of evolution, with algorithmic developments to improve performance along the way.

More recently, specific algorithmic developments have spawned new software. The Stan software package (Carpenter et al., 2017), which has seen extremely rapid adoption in applied fields, implements Hamiltonian Monte Carlo algorithms (see Hoffman & Gelman (2014) and the many references therein). Even more recently, the Blang software package (Bouchard-Côté et al., 2019) uses implementations based on recent research in both SMC and nonreversible MCMC methods. Fortunately, the path proceeding from algorithm research to software implementation to scientific application is one marked by continual upgrades!

### ACKNOWLEDGEMENTS

We thank the guest editors for inviting us to write this article and one anonymous referee for suggestions that improved its initial version. Paul Gustafson is grateful to Harlan Campbell for the continued collaboration and discussion and for leading some of the work described in Section 4. This work was supported by the Natural Sciences and Engineering Research Council of Canada through grants RGPIN-249547, RGPIN-03957, and RGPIN-04142.

### REFERENCES

- Andrieu, C. & Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2), 697–725.
- Andrieu, C. & Vihola, M. (2015). Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability*, 25(2), 1030–1077.
- Baragatti, M. & Pudlo, P. (2014). An overview on approximate Bayesian computation. In *ESAIM: Proceedings*, Vol. 44, EDP Sciences, Les Ulis, France, 291–299.
- Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15(1B), 700–738.

- Bendavid, E., Mulaney, B., Sood, N., Shah, S., Bromley-Dulfano, R., Lai, C., Weissberg, Z. et al. (2021). COVID-19 antibody seroprevalence in Santa Clara county, California. *International Journal of Epidemiology*, 50(2), 410–419.
- Bernton, E., Jacob, P. E., Gerber, M., & Robert, C. P. (2017). Inference in generative models using the Wasserstein distance. arXiv preprint, arXiv:1701.05146.
- Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518), 859–877.
- Bornn, L., Pillai, N., Smith, A., & Woodard, D. (2014). One pseudo-sample is enough in approximate Bayesian computation MCMC. *Biometrika*, 99(1), 1–10.
- Bouchard-Côté, A., Chern, K., Cubranic, D., Hosseini, S., Hume, J., Lepur, M., Ouyang, Z., & Sgarbi, G. (2019). Blang: Bayesian declarative modelling of arbitrary data structures. arXiv preprint, arXiv:1912.10396.
- Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*, CRC Press, Boca Raton, FL.
- Burstyn, I., Goldstein, N. D., & Gustafson, P. (2020). Towards reduction in bias in epidemic curves due to outcome misclassification through Bayesian analysis of time-series of laboratory test results: Case study of COVID-19 in Alberta, Canada and Philadelphia, USA. *BMC Medical Research Methodology*, 20, 1–10.
- Campbell, H., de Valpine, P., Maxwell, L., de Jong, V. M., Debray, T., Jänisch, T., & Gustafson, P. (2021). Bayesian adjustment for preferential testing in estimating the COVID-19 infection fatality rate. *Annals of Applied Statistics*, 16(1), 436–459.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Changye, W. & Robert, C. P. (2019). Parallelising MCMC via random forests. arXiv preprint, arXiv:1911.09698.
- Cook, S. A. (1971). The complexity of theorem-proving procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, ACM, New York, 151–158.
- Cui, T., Peeters, L., Pagendam, D., Pickett, T., Jin, H., Crosbie, R. S., Raiber, M., Rassam, D. W., & Gilfedder, M. (2018). Emulator-enabled approximate Bayesian computation (ABC) and uncertainty analysis for computationally expensive groundwater models. *Journal of Hydrology*, 564, 191–207.
- Drovandi, C. C. (2018). ABC and indirect inference. In *Handbook of Approximate Bayesian Computation*, Chapman & Hall/CRC Press, Boca Raton, FL, 179–209.
- Drovandi, C. C., Grazian, C., Mengersen, K., & Robert, C. (2018). Approximating the likelihood in ABC. In *Handbook of Approximate Bayesian Computation*, Chapman & Hall/CRC Press, Boca Raton, FL, 321–368.
- Entezari, R., Craiu, R. V., & Rosenthal, J. S. (2018). Likelihood inflating sampling algorithm. *Canadian Journal of Statistics*, 46(1), 147–175.
- Fearnhead, P. & Prangle, D. (2012). Constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 419–474.
- Filippi, S., Barnes, C. P., Cornebise, J., & Stumpf, M. P. (2013). On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology*, 12(1), 87–107.
- Gelman, A. & Carpenter, B. (2020). Bayesian analysis of tests with unknown specificity and sensitivity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(5), 1269–1283.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472.
- Gilks, W. R., Thomas, A., & Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1), 169–177.
- Günther, F., Berger, U., Höhle, M., Bender, A., Wildner, M., Heid, I. M., & Küchenhoff, H. (2021). Analysis of COVID-19 case numbers: Adjustment for diagnostic misclassification on the example of German case reporting data. <https://doi.org/10.1101/2020.09.29.20203877>
- Gustafson, P. (2015). *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited data*, CRC Press, Boca Raton, FL.

- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Hoffman, M. D. & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Huggins, J., Campbell, T., & Broderick, T. (2016). Coresets for scalable Bayesian logistic regression. *Advances in Neural Information Processing Systems*, 29, 4080–4088.
- Jones, G. L. & Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, 16(4), 312–334.
- Jones, G. L. & Hobert, J. P. (2004). Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2), 784–817.
- Laskey, K. B. & Myers, J. W. (2003). Population Markov chain Monte Carlo. *Machine Learning*, 50(1-2), 175–196.
- Lee, A. (2012). On the choice of MCMC kernels for approximate Bayesian computation with SMC samplers. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, IEEE, Berlin, Germany, 1–12.
- Lee, A., Andrieu, C., & Doucet, A. (2012). Discussion of constructing summary statistics for approximate Bayesian computation: Semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3), 449–450.
- Levi, E. & Craiu, R. V. (2022). Finding our way in the dark: Approximate MCMC for approximate Bayesian methods. *Bayesian Analysis*, 17(1), 193–221.
- Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25), 3049–3067.
- Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS-A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4), 325–337.
- Manski, C. F. (2003). *Partial Identification of Probability Distributions*, Springer Science & Business Media, Berlin, Germany.
- Marin, J.-M., Pudlo, P., Robert, C. P., & Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, 22(6), 1167–1180.
- Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26), 15324–15328.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092.
- Meyn, S. P. & Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, 4(4), 981–1011.
- Neiswanger, W., Wang, C., & Xing, E. Asymptotically exact, embarrassingly parallel MCMC. arXiv preprint, arXiv:1311.4780.
- Nemeth, C. & Sherlock, C. (2018). Merging MCMC subposteriors through Gaussian-process approximations. *Bayesian Analysis*, 13(2), 507–530.
- Oyebamiji, O. K., Edwards, N. R., Holden, P. B., Garthwaite, P. H., Schaphoff, S., & Gerten, D. (2015). Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15(6), 499–525.
- Plumlee, M., Asher, T. G., Chang, W., & Bilskie, M. V. (2021). High-fidelity hurricane surge forecasting using emulation and sequential experiments. *The Annals of Applied Statistics*, 15(1), 460–480.
- Prangle, D. (2015). Summary statistics in approximate Bayesian computation. arXiv preprint, arXiv:1512.05633.
- Price, L. F., Drovandi, C. C., Lee, A., & Nott, D. J. (2018). Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1), 1–11.
- Qin, Q. & Hobert, J. P. (2017). Asymptotically stable drift and minorization for Markov chains with application to Albert and Chib’s algorithm. arXiv preprint, arXiv:1712.08867.
- Quiroz, M., Kohn, R., Villani, M., & Tran, M.-N. (2018). Speeding up MCMC by efficient data subsampling. *Journal of the American Statistical Association*, 114(526), 831–843. <https://doi.org/10.1080/01621459.2018.1448827>.
- Rajaratnam, B. & Sparks, D. (2015). MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. arXiv preprint, arXiv:1508.00947.
- Roberts, G. O., Gelman, A., & Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1), 110–120.



- Roberts, G. O. & Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1), 255–268.
- Roberts, G. O. & Rosenthal, J. S. (2001). Optimal scaling for various Metropolis–Hastings algorithms. *Statistical Science*, 16(4), 351–367.
- Roberts, G. O. & Rosenthal, J. S. (2016). Complexity bounds for Markov chain Monte Carlo algorithms via diffusion limits. *Journal of Applied Probability*, 53(2), 410–420.
- Roberts, G. O. & Tweedie, R. L. (1999). Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and Their Applications*, 80(2), 211–229.
- Rosenthal, J. S. (1995a). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90(430), 558–566.
- Rosenthal, J. S. (1995b). Rates of convergence for Gibbs sampling for variance components models. *The Annals of Statistics*, 23, 740–761.
- Rosenthal, J. S. (1996). Analysis of the Gibbs sampler for a model related to James–Stein estimators. *Statistics and Computing*, 6(3), 269–275.
- Rosenthal, J. S. (2000). Parallel computing and Monte Carlo algorithms. *Far East Journal of Theoretical Statistics*, 4(2), 207–236.
- Rosenthal, J. S. (2002). Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7, 123–128.
- Rue, H., Riebler, A., Sorbye, S., Illian, J., Simpson, D., & Lindgren, F. (2017). Bayesian computing with INLA: A review. *Annual Review of Statistics and its Application*, 4, 395–421.
- Schervish, M. J. (1988). Applications of parallel computation to statistical inference. *Journal of the American Statistical Association*, 83(404), 976–983.
- Schubert, E. & Gertz, M. (2018). Numerically stable parallel computation of (co-)variance. variance. In *Proceedings of the 30th International Conference on Scientific and Statistical Database Management*, Bozen-Bolzano, Italy, 1–12.
- Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., & McCulloch, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 11(2), 78–88.
- Sisson, S., Fan, Y., & Beaumont, M. (2018a). Overview of approximate Bayesian computation. arXiv preprint, arXiv:1802.09720.
- Sisson, S. A., Fan, Y., & Beaumont, M. (2018b). *Handbook of Approximate Bayesian Computation*, Chapman & Hall/CRC Press, Boca Raton, FL.
- Sisson, S. A., Fan, Y., & Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6), 1760–1765.
- Suchard, M. A., Wang, Q., Chan, C., Frelinger, J., Cron, A., & West, M. (2010). Understanding GPU programming for statistical computation: Studies in massively parallel massive mixtures. *Journal of Computational and Graphical Statistics*, 19(2), 419–438.
- Tawn, N. G., Roberts, G. O., & Rosenthal, J. S. (2020). Skew Brownian motion and complexity of the ALPS algorithm. Under revision.
- Terenin, A., Simpson, D., & Draper, D. (2020). Asynchronous Gibbs sampling. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, PMLR, Palermo, Sicily, Italy, 144–154.
- Thompson, E. (2011). MCMC in the analysis of genetic data on related individuals. In *Handbook of Markov Chain Monte Carlo*, Vol. 339. CRC Press, Boca Raton, FL.
- Wang, X. & Dunson, D. B. (2013). Parallelizing MCMC via Weierstrass sampler. arXiv preprint, arXiv:1312.4605.
- Wilkinson, D. J. (2006). Parallel Bayesian computation. *Statistics Textbooks and Monographs*, 184, 477.
- Wood, S. N. (2010). Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310), 1102.
- Woodard, D. B. & Rosenthal, J. S. (2013). Convergence rate of Markov chain methods for genomic motif discovery. *The Annals of Statistics*, 41, 91–124.
- Woodard, D. B., Schmidler, S. C., & Huber, M. (2009a). Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electronic Journal of Probability*, 14, 780–804.
- Woodard, D. B., Schmidler, S. C., & Huber, M. (2009b). Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *The Annals of Applied Probability*, 19(2), 617–640.



- Wu, S. L., Mertens, A. N., Crider, Y. S., Nguyen, A., Pokpongkiat, N. N., Djajadi, S., Seth, A. et al. (2020). Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, 11(1), 1–10.
- Yang, J., & Rosenthal, J. S. (2017). Complexity results for MCMC derived from quantitative bounds. Under revision.
- Yang, Y., Wainwright, M. J., & Jordan, M. I. (2016). On the computational complexity of high-dimensional Bayesian variable selection. *The Annals of Statistics*, 44(6), 2497–2532.
- 

*Received 8 October 2021*

*Accepted 1 March 2022*