Reflections on Bayesian Inference and Markov Chain Monte Carlo

Radu V. Craiu[1], Paul Gustafson[2], and Jeffrey S. Rosenthal[3]

[1]Department of Statistical Sciences, University of Toronto, CRM-SSC Prize winner (2016), email: radu.craiu@utoronto.ca *

[2]Department of Statistics, University of British Columbia, CRM-SSC Prize winner (2008), SSC Gold Medal winner (2020), email: gustaf@stat.ubc.ca †

[3]Department of Statistical Sciences, University of Toronto, CRM-SSC Prize winner (2006), SSC Gold Medal winner (2013), email: jeff@math.toronto.edu ‡

February 26, 2022

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.

1 INTRODUCTION

On the occasion of the 50-th 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

*Work supported by NSERC funding through RGPIN-249547
†Work supported by NSERC funding through RGPIN-2019-03957
‡Work supported by NSERC funding through RGPIN-2019-04142

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 π. While other applications exist, the most common use of MCMC is in cases where π is a posterior distribution

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

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

Note that the denominator in (1) is, usually, mathematically intractable and impedes the calculation of posterior quantities like

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

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

The most basic version of MCMC is the *Metropolis algorithm* (Metropolis et al., 1953). From a given state θ, it proceeds by first *proposing* to move to a new state ω, and then either *accepting* that proposal (i.e., moving to ω), or *rejecting* that proposal (i.e., staying at θ). 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 ω from θ, as of proposing θ from ω), this procedure ensures that the resulting Markov chain will be reversible with respect to π, and thus have π 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 to the stationary distribution and expected values, thus facilitating sampling and estimation. These and related algorithms have been extremely influential in Bayesian computation and led to many thousands of research papers 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. Rosenthal (1995a,b, 1996, 2002); Meyn and Tweedie (1994); Roberts and Tweedie (1999);

Jones and Hobert (2001, 2004); Baxendale (2005)) and practical (e.g. Gelman and 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 which are perfectly satisfactory in the former context may still fail in the latter context. This issue was driven home to me when I realised to my surprise that one of my best specific quantitative theoretical convergence bounds, proving that a modest 140 iterations sufficed to get within 0.01 of stationarity for a reasonably 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. While various MCMC researchers had considered computational complexity perspectives to some extent in earlier MCMC analysis (e.g. Rosenthal (1995b); Woodard et al. (2009a,b)), this has taken on new prominence in recent years (e.g. Woodard and Rosenthal (2013); Rajaratnam and Sparks (2015); Yang et al. (2016); Qin and Hobert (2017); Yang and Rosenthal (2017); Tawn et al. (2020)). These new results are very promising, but they tend to apply only to specific models, or to require unverifiable assumptions, thus limiting their utility as a general solution to this challenge.

In some cases, the problem of computational complexity of MCMC can be avoided entirely, by instead using an approximation method such as Variational Bayes (Blei et al., 2017) or INLA (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 make use of the results of Roberts and co-authors about optimal scaling of proposal distributions (e.g. Roberts et al. (1997); Roberts and Rosenthal (1998)). They involve speeding up the original algorithm by a fixed power of the dimension d (e.g. d or $d^{1/3}$), and then proving that as $d \to \infty$, the resulting algorithm converges to a diffusion limit whose speed can then be optimised. Since the limiting diffusion no longer depends on dimension, this 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 formalised in Roberts and Rosenthal (2016), where it was proven that traditional random-walk Metropolis algorithms would converge to stationary in $O(d)$ iterations, while Metropolis-adjusted Langevin algorithms would converge in $O(d^{1/3})$ iterations. However, these results were established only under the very strong assumptions which had been required for the original diffusion limits, including assuming that the target distribution factored into i.i.d. components (although in simulations they do appear to approximately hold much more generally (Roberts and Rosenthal, 2001)). Those theoretical assumptions would never hold in practice, leading to questions about how generally the corresponding results will hold in real MCMC applications.

So where does that leave us? We can say without hesitation that there has been a tremendous amount of success using and verifying MCMC on moderately-sized problems. However, when MCMC is used on much larger-scale 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 general 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 lots of 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 computation methods are ineffective, MCMC sampling has been widely adopted due to ease of implementation and available software. Expanding on the discourse of Section 2, let me consider Hastings' generalization of the Metropolis sampler (Hastings, 1970), the so-called Metropolis-Hastings algorithm (henceforth, MH), which is probably the most widely used MCMC sampler.

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

Step 1 A proposal ω_{t+1} is drawn from a proposal density $q(\omega|\theta_t)$;

Step 2 Set

$$\theta_{t+1} = \begin{cases} \omega_{t+1} & \text{with probability } \alpha \\ \theta_t & \text{with probability } 1-\alpha \end{cases}$$

where

$$\alpha = \min\left\{1, \frac{\pi(\omega_{t+1}|\mathbf{y}_0)q(\theta_t|\omega_{t+1})}{\pi(\theta_t|\mathbf{y}_0)q(\omega_{t+1}|\theta_t)}\right\}. \tag{3}$$

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

3.2 Challenges posed by big data

An important initial challenge to the classical MCMC procedures is presented by the sheer increase in data volume. When the likelihood contains 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 (3).

In other areas of statistical computation, algorithm designers have taken advantage of parallelization strategies in which the 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 and Gertz, 2018). Alas, MCMC samplers are notoriously resistant to parallelization, forcing the design and adoption of new ideas, as well as opening the door to different approximation techniques. For instance, Suchard et al. (2010) propose the use of a GPU's multiple processors for speeding up a block Gibbs algorithm. Their approach exploits the GPU's multiple cores which allow parallelization of computing tasks within each MCMC iteration. The proposed approach relies on intrinsic synchronicity of the parallel tasks, i.e. all cores must complete their tasks before the updating of the chain can be completed. 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 and Myers, 2003; Wilkinson, 2006) such that each of the workers 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 outside realistic bounds. It has been recognized that synchronous coordination (workers must wait on each other to finish updates before moving on to

the next step) and frequent communication between workers slow down computation significantly and should be avoided. Some progress on asynchronous updating of Gibbs samplers is reported in Terenin et al. (2020), albeit under relatively stringent conditions.

A strategy that reduces communication between workers and eliminates the need to store all the data on a single server can be achieved via the *embarassingly parallel* design in which the data is partitioned into K equally-sized subsets, called *shards*, with each shard analyzed independently by a different worker. Each worker uses an MCMC algorithm to draw samples from the posterior corresponding to that data shard, which I call a *sub-posterior*. Some essential MCMC-related questions are: 1) which sub-posterior distributions should one build for each shard, and 2) how to combine the MCMC samples obtained from each sub-posterior so that we can recover the information that would have been produced by an MCMC sample from the full posterior distribution. The sub-posteriors developed in the literature for the s-th shard have the form

$$\pi^{(s)} \propto p(\theta)^{a_s} f(\mathbf{y}_0^{(s)}|\theta)^{b_s}, \tag{4}$$

where $\mathbf{y}_0^{(s)}$ is the s-th data shard and $a_s, b_s \in \mathbf{R}$ are user defined. For instance, Scott et al. (2016), Neiswanger et al. (2013) and Wang and Dunson (2013) all use $a_s = 1/K$ and $b_s = 1$, for all $1 \leq s \leq K$. The approaches in combining sub-samples are different. Specifically, Neiswanger et al. (2013) approximate each sub-posterior using kernel density estimators, while Wang and Dunson (2013) use the Weierstrass transformation. Similarly, Nemeth and Sherlock (2018) approximate the sub-posteriors using Gaussian processes in the case $a_s = 1/K, b_s = 1$ for all $1 \leq s \leq K$. Another option is to use $a_s = 1$, $b_s = K$ for all $1 \leq s \leq K$ as in the likelihood inflating algorithm (LISA) that was developed by Entezari et al. (2018) for Bayesian additive regression trees. More recently, Changye and Robert (2019) propose to use $a_s = \lambda_s/K$ and $b_s = \lambda_s$ and the sub-posteriors are first approximated using random forests and then reweighted via importance sampling.

These divide-and-conquer methods are well-justified for Gaussian posteriors and sub-posteriors but theory lags behind once departures from normality are recorded. The most general "fix" is to use importance sampling so that each sub-posterior sample is properly weighted, but this strategy adds significantly to the computational burden. Extending divide-and-conquer techniques to models for non-iid data is also challenging because batches are not independent, or even exchangeable.

Since the size of the data is posing serious challenges, it is reasonable to consider using only a subset of the sample to speed-up the 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. Alas, such simplicity is almost never available in complex models, so new ideas are needed.

Suppose that the log-likelihood is obtained from adding N independent terms, each corresponding to an independent item in the sample,

$$l_N(\theta|\mathbf{y}_0) = \sum_{i=1}^{N} l_i(\theta), \tag{5}$$

where $l_i(\theta) = \log f(y_{0,i}|\theta)$. It is tempting to consider a random subsampling of size r of the data items $\{i_1, \ldots, i_r\}$ and to 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 a MH sampling algorithm, will alter the target distribution of the chain and can potentially lead to large inferential errors. This concern is alleviated if an unbiased estimator of the likelihood is available, due to the pseudo-marginal approach developed by Andrieu and Roberts (2009). They showed that the target distribution of a MH sampling algorithm remains unchanged if the likelihood is replaced by an unbiased estimator when computing the acceptance ratio (3). An important addendum is that the sampling efficiency of the new algorithm degrades as the variance of the ratio of likelihoods increases (Andrieu and Vihola, 2015), thus discouraging a choice of r much smaller than N .

Some of these challenges are tackled by Quiroz et al. (2018) who specify a subsampling-based approach in which a different sub-sample of size r is used at each iteration of the MCMC sampler. They introduce an estimator of the likelihood that is approximatively bias-corrected. In order to keep in check the variance of the likelihood ratio, they introduce control variates. Since the estimator is not exactly unbiased, the transition kernel of the MH chain deviates from the original one and its target distribution is perturbed. However, Quiroz et al. (2018) are able to 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 r.

Keeping with the "data trimming" theme, Huggins et al. (2016) consider a different approach based on the concept of Bayesian coreset. In this case, a single data subset 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 likelihood and the one corresponding to the coreset are derived theoretically. For Gaussian models, Huggins et al. (2016) also derive bounds for the discrepancies 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. The design of large studies will lead not only to large volumes of data, but also sophisticated questions. In turn, the latter can be satisfactorily answered by considering highly complex models that elude analytical formulations available in closed form. Computer emulators for complex phenomena, e.g. the path of hurricanes (Cui et al., 2018; Plumlee et al., 2021) 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: the Approximate Bayesian Computation (ABC) (Marin et al., 2012; Baragatti and Pudlo, 2014; Sisson et al., 2018a; Drovandi, 2018) and the Bayesian Synthetic Likelihood (BSL)(Wood, 2010; Drovandi et al., 2018; Price et al., 2018). Both algorithms are effective when they are combined with Markov chain Monte Carlo sampling schemes to produce samples from an approximation of the posterior.

In its simplest form, the ABC is an accept/reject sampler. Given observed data $\mathbf{y}_0$, a user-defined threshold, $\epsilon > 0$, a distance $d : \mathbf{R}^p \times \mathbf{R}^p \to \mathbf{R}_+$ and 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 θ^* as a sample from the approximate posterior $\pi_\epsilon(\theta|S(\mathbf{y}_0))$, the marginal (in θ) of the joint distribution

$$\pi_\epsilon(\theta, \mathbf{y}|S(\mathbf{y}_0)) \propto p(\theta) f(\mathbf{y}|\theta) \mathbf{1}_{\{d(S(\mathbf{y}), S(\mathbf{y}_0)) < \epsilon\}}. \tag{6}$$

Note that 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 θ and $\mathbf{y}$ take discrete values. Then, one can easily see that

$$\Pr(\theta = \theta_0|S(\mathbf{y}_0)) \propto p(\theta_0) \Pr(S(\mathbf{y}) = S(\mathbf{y}_0)|\theta = \theta_0) \propto \pi(\theta_0|\mathbf{y}_0), \tag{7}$$

where (7) 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 papers and ideas that simply cannot all 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, ϵ or S. Theory-backed recommendations for the choice of S can be found in Fearnhead and 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 the model parameters, the prior and posterior distributions may have non-overlapping regions with significant mass. Hence, parameter values that are drawn from the prior, as in S1, will be rarely retained. Recognizing this, Marjoram et al. (2003) proposed an ABC-MCMC algorithm which relies on building a Metropolis-Hastings (MH) transition kernel, with state space $\{(\theta, \mathbf{y}) \in \mathbf{R}^q \times \mathcal{X}^n\}$, proposal distribution at iteration t, $q(\theta|\theta_t) \times f(\mathbf{y}|\theta)$, and the target $\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0)$ given in (6). Note that the acceptance probability (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's sampler, motivated by low acceptance probabilities for small values of ϵ. For instance, Lee et al. (2012) note that

$$\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0) \propto P(d(S(\mathbf{y}), S(\mathbf{y}_0)) < \epsilon|\theta) \tag{8}$$

and thus can replace $\pi_\epsilon(\theta, \mathbf{y}|\mathbf{y}_0)$ in the calculation of the acceptance probability with one of its unbiased estimators, $J^{-1} \sum_{j=1}^{J} \mathbf{1}_{\{d(S(\mathbf{y}_j), S(\mathbf{y}_0)) < \epsilon\}}$ where $J \geq 1$ and each $\mathbf{y}_j$ is independently simulated from $f(\mathbf{y}|\theta)$. The acceptance probability tends to increase, 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 (8) when ϵ 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 et al., 2007; Lee, 2012; Filippi et al., 2013) and rely on a user-specified decreasing sequence $\epsilon_0 > \cdots > \epsilon_J$. A comprehensive coverage of ABC-related theory and computational techniques can be found in Sisson et al. (2018b) and references therein.

An alternative approach to bypass the intractability of the sampling distribution is proposed by Wood (2010). His approach is based on the working assumption that the conditional distribution for a user-defined statistic $S(\mathbf{y})$ given θ is Gaussian with mean μ_θ and covariance matrix Σ_θ. The *Synthetic Likelihood* (SL) procedure assigns to each θ the likelihood $SL(\theta) = \mathcal{N}(s_0; \mu_\theta, \Sigma_\theta)$, where $s_0 = S(\mathbf{y}_0)$ and $\mathcal{N}(x; \mu, \Sigma)$ denotes the density of a normal with mean μ and covariance Σ. SL can be used for maximum likelihood estimation as in Wood (2010) or within the Bayesian paradigm as proposed by Drovandi et al. (2018) and Price et al. (2018). The latter work proposes to sample the approximate posterior generated by the Bayesian Synthetic Likelihood (BSL) approach, $\pi(\theta|s_0) \propto p(\theta)\mathcal{N}(s_0; \mu_\theta, \Sigma_\theta)$, using a MH sampler. Direct calculation of the acceptance

probability is not possible because the conditional mean and covariance are unknown for any θ. However, both can be estimated based on m statistics $(s_1, \cdots, s_m)$ sampled from their conditional distribution given θ. More precisely, after simulating $\mathbf{y}_i \sim f(\mathbf{y}|\theta)$ and setting $s_i = S(\mathbf{y}_i)$, $i = 1, \cdots, m$, one can estimate

$$\begin{aligned} \hat{\mu}_\theta &= \frac{\sum_{i=1}^m s_i}{m}, \\ \hat{\Sigma}_\theta &= \frac{\sum_{i=1}^m (s_i - \hat{\mu}_\theta)(s_i - \hat{\mu}_\theta)^T}{m-1}, \end{aligned} \tag{9}$$

so that the synthetic likelihood is

$$SL(\theta|\mathbf{y}_0) = \mathcal{N}(S(\mathbf{y}_0); \hat{\mu}_\theta, \hat{\Sigma}_\theta). \tag{10}$$

A MH algorithm designed to sample from the posterior $\pi_{SL}(\theta|\mathbf{y}_0) \propto p(\theta) SL(\theta|\mathbf{y}_0)$ will require a proposal distribution $q(\theta^*|\theta)$. At step t the underlying Monte Carlo chain is updated using:

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

SL2 Sample $\mathbf{y}_1, \ldots, \mathbf{y}_m \overset{iid}{\sim} f(\mathbf{y}|\theta^*)$ and compute μ_{θ^*} and Σ_{θ^*} as in (9) to obtain $SL(\theta^*|\mathbf{y}_0)$

SL3 Set $\theta_{t+1} = \theta^*$ with probability $\alpha_t = \min\left\{1, \frac{q(\theta_t|\theta^*)\pi_{SL}(\theta^*|\mathbf{y}_0)}{q(\theta^*|\theta_t)\pi_{SL}(\theta_t|\mathbf{y}_0)}\right\}$, and $\theta_{t+1} = \theta_t$ otherwise.

Running MCMC samplers for either ABC or BSL involves generating multiple pseudo-samples and can become extremely costly in situations in which data is 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 and Craiu (2022) propose strategies to minimize the number of pseudo-data simulations. To this end, when computing the MH acceptance ratios, they recommend to reuse 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 produces a perturbation of the target distribution. Their theoretical developments demonstrate that the error can be controlled in the case of independent Metropolis samplers.

3.4 Conclusion

The challenges posed to Bayesian computation are important and require significant reframing of classical methods. We have discussed here a number of methods that address mainly two type of challenges: big data and intractable likelihoods. The two

challenges cannot always be neatly separated and, in fact, I do 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 et al., 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 the methods discussed so far.

Clearly, our efforts to expand the toolbox for Bayesian computation in the modern era are just revving up. Much remains to be done, from devising new conceptual ideas to efficient and automatic implementations. I will 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 controllable and practically controlled under the scenarios 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. And I 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. Frighteningly, the global pandemic has upended society. Interestingly, Bayesian methods have played useful roles within the scientific response to the pandemic. I briefly describe several of these roles, drawing on some work I have been involved in, and some work 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, such that the manufacturer states the test's performance characteristics as part of the "package insert." This would be expressed as values for *sensitivity* and *specificity*, the chance 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, the first serological tests for prior infection were being used in scientific studies (so-called "sero-surveys") by April 2020. In both cases, and because of the necessary haste, testing data were being collected and analyzed with less than typical understanding of the diagnostic test performance. Thus efforts were made to acknowledge the uncertainty in the performance characteristics, a task for which Bayesian methods are well, if not singularly, suited.

To start with the PCR test, Burstyn et al. (2020) presented an early instance of an analysis acknowledging test imperfection. (The first version of the manuscript was posted to `medrxiv.org` on April 11, 2020.) They 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 amongst those receiving a test, as distinct from the daily number of reported positives.

In formulating this Bayesian model, an important design choice was 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 on this point, with both Burstyn et al. (2020) and Günther et al. (2021) having pertinent discussion. It is well understood that the 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, the effective sensitivity will be lower amongst recently infected individuals, compared to those with more established infections. Extrapolating further, when testing a largely asymptomatic population, true positives therein will tend to be recently infected, hence sensitivity will be lower. But if only those with respiratory symptoms are eligible for a test, then the true positives therein will tend to have more established infections, resulting in a higher sensitivity. In fact then, it is 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 level of 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, a link to (a lack of) Bayesian analysis and controversy in both scientific and lay circles arose in mid-April 2020, with an early sero-survey conducted in Santa Clara, California. The initial preprint version of Bendavid et al. (2021) attracted much attention and criticism. One area of concern was as follows. In brief, and in the abstract, consider a diagnostic test for an infection that is rare in

the study population. And say package-insert values of test sensitivity and specificity, $\tilde{Sn}$ and $\tilde{Sp}$, are inputs to an analysis inferring population prevalence of ever-infection from a sample. If the infection is truly quite rare then, even if $\tilde{Sp}$ is very close to one, the estimated prevalence can vary strongly under a small perturbation to $\tilde{Sp}$. Coupled with the urgent need to role out tests and studies at warp-speed, so that "validation studies" to determine diagnostic test properties were based on relatively few (known negative and known positive) test specimens, a problem ensued.

Gelman and Carpenter (2020) reviewed this challenging situation and provided a careful Bayesian analysis acknowledging the uncertainty in test specificity (and sensitivity). They demonstrated much greater *a posteriori* uncertainty about the ever-infected proportion in the study population, relative to the initial version of Bendavid et al. (2021). They also illustrated other modelling features that can be brought to bear quite simply in the 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 prior to the availability of serological tests for past infection, reasonable estimation of the denominator was challenging.

Wu et al. (2020) presented a Bayesian analysis targeting the total number of infections in the U.S., up to mid-April 2020. As part of their 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. Secondarily, and in common with Gelman and Carpenter (2020), they also use a prior distribution to adjust for the imperfect diagnostic test.

The principal finding of Wu et al. (2020) is 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 and 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 a sharper inference would not be justified.

In a related vein, Campbell et al. (2021) developed a Bayesian model admitting

both surveillance data (those tested for current infection choose to be tested, within the confines and recommendations of public health surveillance at the time and place in question) and sero-survey data (a random sample from the target time and place are tested for ever-infection). The surveillance data are 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. A hierarchical prior distribution is ascribed to these jurisdiction-specific preferential testing parameters. Applying this methodology within the confines of an evidence synthesis for European countries in Spring 2020, the infection fatality rate is estimated by a posterior median of 0.53%, 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 on making defensible model and prior assertions, these can be encoded in a Bayesian software package (see Section 5). Then, upon presentation of data, posterior inference ensues. Scientifically, this can be the end of the narrative arc. One has general assurance that reported estimates and posterior uncertainties have been arrived at in a principled way, based on the combined information content of the data and the supplied model and prior assertions.

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

It turns out that the explanation for the above findings arises from consideration of the *partial identification* that underlies the Bayesian model. The concept of partial identification is long-studied, primarily through a frequentist lens, and often 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). At essence, in lower information content 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 which are. By elucidating these inequalities, one can understand directly how the Alberta data say so little about the test sensitivity, and so much about the specificity.

As it happens, such 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 extents of preferential sampling across jurisdictions, the surveillance data will contribute less or more 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. Due in large part to 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 the enabling works. There are longstanding efforts to build software permitting users to compute posterior quantities, without having to be experts in the algorithms. While perhaps we aren't fully there yet, the ideal is that an applied user need just declare model and prior specifications, input the data, and press the "compute posterior" button.

By no means do we attempt to mention all software for Bayesian inference here. We note though that the BUGS (Bayesian inference using Gibbs sampling) software

package (Gilks et al., 1994; Lunn et al., 2000, 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 improving the software 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 work, implements Hamiltonian Monte Carlo algorithms (see Hoffman et al. (2014) and the many references therein). And even more recently, the Blang software package (Bouchard-Côté et al., 2019) uses implementations based on recent research in both sequential Monte Carlo algorithms and non-reversible MCMC methods. Happily, the path proceeding from algorithmic research to software implementation to scientific applications is one marked by continual upgrades!

Acknowledgement

We thank the guest editors for inviting us to write this paper and one anonymous referee for suggestions that improved its initial version. PG is grateful to Harlan Campbell for continued collaboration and discussion, and for leading some of the work described in Section 4.

References

C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725, 2009.

C. Andrieu and M. Vihola. Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms. *The Annals of Applied Probability*, 25(2):1030–1077, 2015.

M. Baragatti and P. Pudlo. An overview on approximate Bayesian computation. In *ESAIM: Proceedings*, volume 44, pages 291–299. EDP Sciences, 2014.

P. H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *The Annals of Applied Probability*, 15(1B):700–738, 2005.

E. Bendavid, B. Mulaney, N. Sood, S. Shah, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra-Walker, J. Tedrow, A. Bogan, et al. Covid-19 antibody seroprevalence in Santa Clara County, California. *International Journal of Epidemiology*, 50(2): 410–419, 2021.

E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert. Inference in generative models using the Wasserstein distance. *arXiv preprint arXiv:1701.05146*, 1(8):9, 2017.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

L. Bornn, N. Pillai, A. Smith, and D. Woodard. One pseudo-sample is enough in approximate Bayesian computation MCMC. *Biometrika*, 99(1):1–10, 2014.

A. Bouchard-Côté, K. Chern, D. Cubranic, S. Hosseini, J. Hume, M. Lepur, Z. Ouyang, and G. Sgarbi. Blang: Bayesian declarative modelling of arbitrary data structures. *arXiv preprint arXiv:1912.10396*, 2019.

S. Brooks, A. Gelman, G. Jones, and X.-L. Meng. *Handbook of Markov chain Monte Carlo*. CRC Press, 2011.

I. Burstyn, N. D. Goldstein, and P. Gustafson. 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, 2020.

H. Campbell, P. de Valpine, L. Maxwell, V. M. de Jong, T. Debray, T. Jänisch, and P. Gustafson. Bayesian adjustment for preferential testing in estimating the COVID-19 infection fatality rate. *Annals of Applied Statistics*, 2021.

B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1):1–32, 2017.

W. Changye and C. P. Robert. Parallelising MCMC via random forests. *arXiv preprint arXiv:1911.09698*, 2019.

S. A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM, 1971.

T. Cui, L. Peeters, D. Pagendam, T. Pickett, H. Jin, R. S. Crosbie, M. Raiber, D. W. Rassam, and M. Gilfedder. Emulator-enabled approximate Bayesian computation (ABC) and uncertainty analysis for computationally expensive groundwater models. *Journal of Hydrology*, 564:191–207, 2018.

C. C. Drovandi. ABC and indirect inference. *Handbook of Approximate Bayesian Computation*, pages 179–209, 2018.

C. C. Drovandi, C. Grazian, K. Mengersen, and C. Robert. Approximating the likelihood in ABC. In *Handbook of Approximate Bayesian Computation*, pages 321–368. Chapman and Hall/CRC, 2018.

R. Entezari, R. V. Craiu, and J. S. Rosenthal. Likelihood inflating sampling algorithm. *Canadian Journal of Statistics*, 46(1):147–175, 2018.

P. Fearnhead and D. Prangle. 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, 2012.

S. Filippi, C. P. Barnes, J. Cornebise, and M. P. Stumpf. On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo. *Statistical Applications in Genetics and Molecular Biology*, 12(1):87–107, 2013.

A. Gelman and B. Carpenter. Bayesian analysis of tests with unknown specificity and sensitivity. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69 (5):1269–1283, 2020.

A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, pages 457–472, 1992.

W. R. Gilks, A. Thomas, and D. J. Spiegelhalter. A language and program for complex Bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1):169–177, 1994.

F. Günther, U. Berger, M. Höhle, A. Bender, M. Wildner, I. M. Heid, and H. Küchenhoff. Analysis of COVID-19 case numbers: Adjustment for diagnostic misclassification on the example of German case reporting data. *medRxiv*, 2021.

P. Gustafson. *Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited data*. CRC Press, 2015.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

M. D. Hoffman, A. Gelman, et al. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

J. Huggins, T. Campbell, and T. Broderick. Coresets for scalable Bayesian logistic regression. *Advances in Neural Information Processing Systems*, 29:4080–4088, 2016.

G. L. Jones and J. P. Hobert. Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, pages 312–334, 2001.

G. L. Jones and J. P. Hobert. Sufficient burn-in for Gibbs samplers for a hierarchical random effects model. *The Annals of Statistics*, 32(2):784–817, 2004.

K. B. Laskey and J. W. Myers. Population Markov chain Monte Carlo. *Machine Learning*, 50(1-2):175–196, 2003.

A. Lee. On the choice of MCMC kernels for approximate Bayesian computation with SMC samplers. In *Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages 1–12. IEEE, 2012.

A. Lee, C. Andrieu, and A. Doucet. Discussion of constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *JR Stat. Soc. Ser. B Stat. Methodol*, 74(3):449–450, 2012.

E. Levi and R. V. Craiu. Finding our way in the dark: Approximate MCMC for approximate Bayesian methods. *Bayesian Analysis*, 17(1):193–221, 2022.

D. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067, 2009.

D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS-a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4): 325–337, 2000.

C. F. Manski. *Partial Identification of Probability Distributions*. Springer Science & Business Media, 2003.

J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328, 2003.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

S. P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, pages 981–1011, 1994.

W. Neiswanger, C. Wang, and E. Xing. Asymptotically exact, embarrassingly parallel MCMC. *arXiv preprint arXiv:1311.4780*, 2013.

C. Nemeth and C. Sherlock. Merging MCMC subposteriors through Gaussian-process approximations. *Bayesian Analysis*, 13(2):507–530, 2018.

O. K. Oyebamiji, N. R. Edwards, P. B. Holden, P. H. Garthwaite, S. Schaphoff, and D. Gerten. Emulating global climate change impacts on crop yields. *Statistical Modelling*, 15(6):499–525, 2015.

M. Plumlee, T. G. Asher, W. Chang, and M. V. Bilskie. High-fidelity hurricane surge forecasting using emulation and sequential experiments. *The Annals of Applied Statistics*, 15(1):460–480, 2021.

D. Prangle. Summary statistics in approximate bayesian computation. *arXiv preprint arXiv:1512.05633*, 2015.

L. F. Price, C. C. Drovandi, A. Lee, and D. J. Nott. Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics*, 27(1):1–11, 2018.

Q. Qin and J. P. Hobert. Asymptotically stable drift and minorization for Markov chains with application to Albert and Chib's algorithm. *arXiv preprint 1712.08867*, 2017.

M. Quiroz, R. Kohn, M. Villani, and M.-N. Tran. Speeding up MCMC by efficient data subsampling. *Journal of the American Statistical Association*, 2018.

B. Rajaratnam and D. Sparks. MCMC-based inference in the era of big data: A fundamental analysis of the convergence complexity of high-dimensional chains. *arXiv preprint 1508.00947*, 2015.

G. O. Roberts and J. S. Rosenthal. Optimal scaling of discrete approximations to Langevin diffusions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):255–268, 1998.

G. O. Roberts and J. S. Rosenthal. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16(4):351–367, 2001.

G. O. Roberts and J. S. Rosenthal. Complexity bounds for Markov chain Monte Carlo algorithms via diffusion limits. *Journal of Applied Probability*, 53(2):410–420, 2016.

G. O. Roberts and R. L. Tweedie. Bounds on regeneration times and convergence rates for Markov chains. *Stochastic Processes and their Applications*, 80(2):211–229, 1999.

G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

J. S. Rosenthal. Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566, 1995a.

J. S. Rosenthal. Rates of convergence for Gibbs sampling for variance components models. *The Annals of Statistics*, 23:740–761, 1995b.

J. S. Rosenthal. Analysis of the Gibbs sampler for a model related to James-Stein estimators. *Statistics and Computing*, 6(3):269–275, 1996.

J. S. Rosenthal. Parallel computing and Monte Carlo algorithms. *Far East Journal of Theoretical Statistics*, 4(2):207–236, 2000.

J. S. Rosenthal. Quantitative convergence rates of Markov chains: A simple account. *Electronic Communications in Probability*, 7:123–128, 2002.

H. Rue, A. Riebler, S. Sorbye, J. Illian, D. Simpson, and F. Lindgren. Bayesian computing with INLA: A review. *Annual Review of Statistics and its Application*, 4: 395–421, 2017.

M. J. Schervish. Applications of parallel computation to statistical inference. *Journal of the American Statistical Association*, 83(404):976–983, 1988.

E. Schubert and M. Gertz. Numerically stable parallel computation of (co-) variance. In *Proceedings of the 30th International Conference on Scientific and Statistical Database Management*, pages 1–12, 2018.

S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, and R. E. McCulloch. Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management*, 11(2):78–88, 2016.

S. Sisson, Y. Fan, and M. Beaumont. Overview of approximate Bayesian computation. *arXiv preprint arXiv:1802.09720*, 2018a.

S. A. Sisson, Y. Fan, and M. M. Tanaka. Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765, 2007.

S. A. Sisson, Y. Fan, and M. Beaumont. *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, 2018b.

M. A. Suchard, Q. Wang, C. Chan, J. Frelinger, A. Cron, and M. West. Understanding GPU programming for statistical computation: Studies in massively parallel massive mixtures. *Journal of Computational and Graphical Statistics*, 19(2):419–438, 2010.

N. G. Tawn, G. O. Roberts, and J. S. Rosenthal. Skew Brownian Motion and Complexity of the ALPS Algorithm. Under revision, 2020.

A. Terenin, D. Simpson, and D. Draper. Asynchronous Gibbs sampling. In *International Conference on Artificial Intelligence and Statistics*, pages 144–154. PMLR, 2020.

E. Thompson. MCMC in the analysis of genetic data on related individuals. *Handbook of Markov Chain Monte Carlo*, page 339, 2011.

X. Wang and D. B. Dunson. Parallelizing MCMC via Weierstrass sampler. *arXiv preprint arXiv:1312.4605*, 2013.

D. J. Wilkinson. Parallel Bayesian computation. *Statistics Textbooks and Monographs*, 184:477, 2006.

S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102, 2010.

D. B. Woodard and J. S. Rosenthal. Convergence rate of Markov chain methods for genomic motif discovery. *The Annals of Statistics*, 41:91–124, 2013.

D. B. Woodard, S. C. Schmidler, and M. Huber. Sufficient conditions for torpid mixing of parallel and simulated tempering. *Electronic Journal of Probability*, 14:780–804, 2009a.

D. B. Woodard, S. C. Schmidler, and M. Huber. Conditions for rapid mixing of parallel and simulated tempering on multimodal distributions. *The Annals of Applied Probability*, pages 617–640, 2009b.

S. L. Wu, A. N. Mertens, Y. S. Crider, A. Nguyen, N. N. Pokpongkiat, S. Djajadi, A. Seth, M. S. Hsiang, J. M. Colford, A. Reingold, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications*, 11(1):1–10, 2020.

J. Yang and J. S. Rosenthal. Complexity results for MCMC derived from quantitative bounds. Under revision, 2017.

Y. Yang, M. J. Wainwright, and M. I. Jordan. On the computational complexity of high-dimensional Bayesian variable selection. *The Annals of Statistics*, 44(6):2497–2532, 2016.