Barker's algorithm for Bayesian inference with intractable likelihoods

Flávio B. Gonçalves[a], Krzysztof Łatuszyński[b,c] and Gareth O. Roberts[b]

[a]*Departamento de Estatística, Universidade Federal de Minas Gerais*
[b]*Department of Statistics, University of Warwick*
[c]*Alan Turing Institute*

Abstract. In this expository paper, we abstract and describe a simple MCMC scheme for sampling from intractable target densities. The approach has been introduced in Gonçalves, Łatuszyński and Roberts (2017a) in the specific context of jump-diffusions, and is based on the Barker's algorithm paired with a simple Bernoulli factory type scheme, the so called *2-coin algorithm*. In many settings, it is an alternative to standard Metropolis–Hastings pseudo-marginal method for simulating from intractable target densities. Although Barker's is well known to be slightly less efficient than Metropolis–Hastings, the key advantage of our approach is that it allows to implement the "marginal Barker's" instead of the extended state space pseudo-marginal Metropolis–Hastings, owing to the special form of the accept/reject probability. We shall illustrate our methodology in the context of Bayesian inference for discretely observed Wright–Fisher family of diffusions.

1 Introduction

Modern data science is awash with problems with intractable likelihoods, that is, problems in which pointwise evaluation of the likelihood function is either impossible or extremely computationally expensive. Intractability can be caused by data missingness, model complexity, or the sheer size of the data set. Within this context, Bayesian inference is particularly challenging, as its algorithmic workhorse, MCMC, requires large numbers (typically many thousands) of likelihood evaluations, and therefore runs the risk of being prohibitively slow.

We place our problem in the following generic context. Suppose that $\pi(\theta) = \pi(\theta|y)$ is a target (posterior) density (with respect to some dominating measure ν) of parameter θ on state space $\mathcal{X}$ given data set y. Considering the problem of designing an MCMC algorithm that targets π, we shall use $q(\theta, \cdot)$ to denote the density (also with respect to ν) of the proposed transition from θ. The standard Metropolis–Hastings algorithm proposes a move from θ to ϕ accepting with prob-

ability

$$\alpha_{\mathrm{MH}}(\theta, \phi) = 1 \wedge \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \tag{1}$$

requiring at least a function evaluation of $\pi(\phi)$.

Motivated by this, in recent years there has been increased interesting *retrospective* simulation techniques which attempt to simulate from an event of probability (1) directly, and without recourse to calculating the probability itself, see e.g. Beskos, Papaspiliopoulos and Roberts (2008), Beskos et al. (2006b), Papaspiliopoulos and Roberts (2008). These methods rely heavily on being able to simulate from events such as a duplicated data set, y' say, conditional on the parameter ϕ which might have probability (proportional to) $\pi(\phi)$.

However in the current context, even if retrospective simulation of events of probability $\pi(\theta)$ and $\pi(\phi)$ is possible and efficient, that does not directly lead to a solution to the problem of simulation from an event of probability (1) which is a nonlinear expression of $\pi(\theta)$ and $\pi(\phi)$. This problem falls into the category of the classical computational probability problem known as the *Bernoulli factory problem*: given the probability p which we cannot evaluate but where events of this probability can be simulated, how can one simulate from an event of probability $f(p)$? (cf. Von Neumann (1951), Keane and O'Brien (1994)). In the context of (1), the function f takes the form $f(p_1, p_2) = 1 \wedge \frac{c_1 p_1}{c_2 p_2}$ for suitable constants c_1 and c_2, and in many modelling contexts of interest can be equivalently rewritten as

$$f(p_3) = 1 \wedge c_3 p_3. \tag{2}$$

While substantial progress has been made on efficient solutions to the Bernoulli factory problem in recent years (e.g., Nacu and Peres (2005), Łatuszyński et al. (2011) and in more specialised settings Flegal and Herbei (2012), Herbei and Berliner (2014), Huber (2015, 2016)), it is known that in general there does not exist a solution when f takes the form given in (2), cf. Asmussen, Glynn and Thorisson (1992), Łatuszyński et al. (2011). Fortunately, the flexibility of the Metropolis–Hastings algorithm allows us to circumvent this difficulty.

In this paper, we will abstract and describe a framework for the implementation of MCMC with intractable likelihoods by using an alternative acceptance probability to that in (1); namely that introduced in Barker (1965). Although less efficient than (1), the Barker acceptance probability can be simulated using an efficient and simple Bernoulli factory type algorithm we will describe, called the *2-coin algorithm*. The approach has been first used in the complex setting of exact fully Bayesian inference for stochastic differential equations with jumps (Gonçalves, Łatuszyński and Roberts (2017a)) and then also for diffusions with switching regimes (Łatuszyński, Palczewski and Roberts (2017)) and diffusion driven Cox processes (Gonçalves, Roberts and Łatuszyński (2017b)). Here, for expository purposes we will illustrate our algorithm with one toy example and then move on to describe a more realistic example of Bayesian inference for discretely observed

Wright–Fisher diffusions for which the approach of Beskos et al. (2006b) is not applicable.

2 Barker's alternative to Metropolis–Hastings

Peskun (1973) demonstrated that given a proposal kernel q, there are many different choices of acceptance probability $\alpha(\theta, \phi)$ which create a Markov chain with stationary distribution π, going on to show that the Metropolis–Hastings option $\alpha_{\mathrm{MH}}(\theta, \phi)$ in (1) maximises the acceptance probability for all possible transitions $\theta \to \phi$. This in turn implies (via the celebrated Peskun ordering) that the Metropolis–Hastings choice of acceptance probability minimises the asymptotic Monte Carlo variance for estimating integrals of functions in $L^2(\pi)$ (see also Mira and Geyer (1999), Mira (2001); here $L^2(\pi)$ denotes the Hilbert space of functions square integrable with respect to π, for background on functional analytic perspective on Markov chains we refer to Roberts and Rosenthal (1997) and references therein). However, there are many other acceptance probability solutions which can be shown to be almost as good. We will focus on one introduced in Barker (1965):

$$\alpha_B(\theta, \phi) = \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi) + \pi(\phi)q(\phi, \theta)}. \tag{3}$$

It is straightforward to verify that

$$\frac{\alpha_{\mathrm{MH}}(\theta, \phi)}{2} \leq \alpha_B(\theta, \phi) \leq \alpha_{\mathrm{MH}}(\theta, \phi),$$

which indicates that the two algorithms have similar performance. More precisely, in terms of comparing the CLT and its asymptotic variance, the following result holds, which demonstrates that roughly speaking Barker's method is at worst half as good as Metropolis–Hastings.

Proposition 1 (Łatuszyński and Roberts (2013), Theorem 4(ii)). *Let $f \in L^2(\pi)$ and denote the i.i.d. Monte Carlo variance by $\sigma_\pi^2 := \mathrm{Var}_\pi(f)$. If a square root central limit theorem holds for f and the Metropolis–Hastings chain with the CLT assymptotic variance σ_{MH}^2, i.e.*

$$\frac{\sum_{i=1}^N f(\theta_i) - \pi(f)}{\sqrt{N}} \Rightarrow N(0, \sigma_{\mathrm{MH}}^2),$$

then a corresponding CLT holds for f and the Barker chain with CLT asymptotic variance σ_B^2 satisfying

$$\sigma_{\mathrm{MH}}^2 \leq \sigma_B^2 \leq 2\sigma_{\mathrm{MH}}^2 + \sigma_\pi^2.$$

On the other hand, when we can simulate efficiently from events proportional to $\pi(\theta)$, the following *2-coin algorithm* gives a simple way of simulating from an event of probability $\alpha_B(\theta, \phi)$, which we write as $\frac{c_1 p_1}{c_1 p_1 + c_2 p_2}$:

The 2-coin algorithm *for sampling Barker's acceptance probability*:

1. Sample $C_1 \sim \text{Ber}(\frac{c_1}{c_1 + c_2})$;
2. if $C_1 = 1$, sample $C_2 \sim \text{Ber}(p_1)$;

   - if $C_2 = 1$, output 1;
   - if $C_2 = 0$, go back to 1;

3. if $C_1 = 0$, sample $C_2 \sim \text{Ber}(p_2)$;

   - if $C_2 = 1$, output 0;
   - if $C_2 = 0$, go back to 1.

Elementary conditional probability calculations verify that the above *2-coin algorithm* outputs 1 with probability $\frac{c_1 p_1}{c_1 p_1 + c_2 p_2}$ and 0 with probability $\frac{c_2 p_2}{c_1 p_1 + c_2 p_2}$. Furthermore, the number of loops needed until the algorithm stops is distributed as $\text{Geom}(\frac{c_1 p_1 + c_2 p_2}{c_1 + c_2})$ and hence the mean execution time is proportional to $\frac{c_1 + c_2}{c_1 p_1 + c_2 p_2}$.

The approach described in this paper can be seen as an alternative to pseudo-marginal MCMC (Beaumont (2003), Andrieu and Roberts (2009)) in settings where simulation of an event of probability proportional to $\pi(\theta)$ is possible, hence in particular when there is an unbiased positive and bounded estimator of $\pi(\theta)$ (cf. Lemma 2.1 and Theorem 2.7 of Łatuszyński et al. (2011) and also Jacob and Thiery (2015) for a related discussion). While both methods have their advantages, we see that the Barker approach advocated here has the clear distinction of constructing the marginal algorithm, i.e. a Markov chain on the state space $\mathcal{X}$ without any need for additional auxiliary variables, which slow down pseudo-marginal methods in terms of the CLT asymptotic variance (see Andrieu and Vihola (2015), Theorem 7). This paper will not attempt a systematic numerical comparison between these approaches.

3 A simple example

To illustrate how an intractable posterior density π may be written as a product of a known function and an unknown probability from which one can simulate, so that it can be readily sampled via the above Barker's algorithm with 2-coin acceptance, consider the following simple (toy) example. Suppose that (perhaps due to missing data) we can write the posterior distribution π as

$$\pi(\theta) = \mathbf{E}_{\eta \sim h}\big[\pi(\theta \mid \eta)\big]$$

that is, we only have an explicit expression for π as a mixture of conditional densities $\pi(\cdot \mid \eta)$ with known mixing measure h.
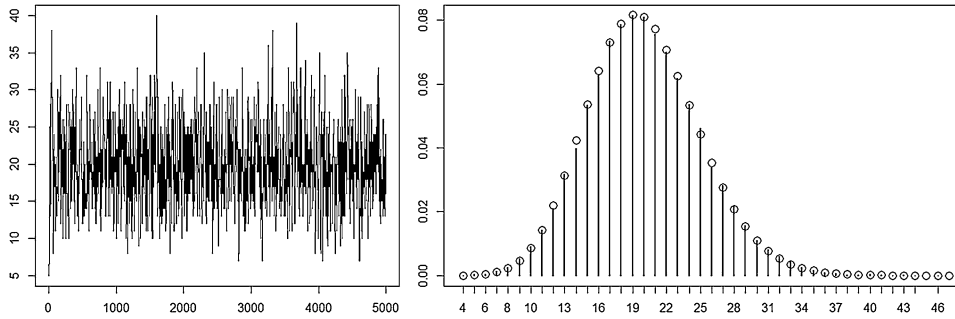
Figure 1 *Trace plot for the first* 5 *thousand iterations and empirical* (*lines*) *and exact* (*circles*) *distribution* $\pi(\theta)$. *The estimated mean and variance are* 20.012 *and* 23.989.

Now, we shall be even more explicit, and assume that $\theta \mid \eta \sim \text{Poisson}(\eta)$ so that

$$\pi(\theta \mid \eta) = \frac{e^{-\eta}\eta^{\theta}}{\theta!} \leq \frac{e^{-\theta}\theta^{\theta}}{\theta!} := d(\theta).$$

Writing $\pi(\theta) = d(\theta)\frac{\pi(\theta)}{d(\theta)}$ yields the desired form since events of probability $p(\theta) := \pi(\theta)/d(\theta)$ are easy to simulate as follows. First, simulate $\eta \sim h$ and an independent standard uniform random variable U. Then the event $\{U \leq \pi(\theta \mid \eta)/d(\theta)\}$ is easily seen to have the desired probability.

Now consider a Barker algorithm which proposes from a symmetric random walk move (i.e. $q(\theta, \phi) = q(\phi, \theta)$). We are thus required to accept the proposed move with probability

$$\alpha_B(\theta, \phi) = \frac{d(\phi)p(\phi)}{d(\phi)p(\phi) + d(\theta)p(\theta)},$$

which we can do using the 2-coin algorithm.

We implement this example with $\eta \sim \text{Gamma}(100, 5)$ and a uniform proposal distribution on $\{\theta - 10, \theta - 9, \ldots, \theta - 1, \theta + 1, \ldots, \theta + 9, \theta + 10\}$, which leads to a 0.367 acceptance rate. The exact distribution $\pi(\theta)$ is Negative Binomial with mean 20 and variance 24. We run the chain for 2.10^6 iterations. Results are reported in Figure 1. The average number of loops in the 2-coin algorithm is 4.7.

4 Barker's MCMC for exact inference in the Wright–Fisher family of diffusions

In this section, we sketch the MCMC methodology for exact Bayesian inference for diffusions introduced in Beskos et al. (2006b) and further developed in Sermaidis et al. (2013). We then consider inference for the Wright–Fisher diffusion and explain why it falls outside the scope of this methodology. Next, for the Wright–Fisher family of diffusions, we present how to design the Barker's based

MCMC algorithm, originally proposed in Gonçalves, Łatuszyński and Roberts (2017a) in the context of jump-diffusions. The algorithm performs exact Bayesian inference by sampling from a Markov chain that has the exact posterior distribution of the parameters and missing paths of the diffusion as its invariant distribution.

4.1 Exact Bayesian inference for discretely observed diffusions

The model considered in this section is that of a stochastic process $Y := \{Y_s : 0 \leq s \leq t\}$ that solves the following SDE parametrised by θ:

$$dY_s = b(Y_s; \theta)\, ds + \sigma(Y_s; \theta)\, dW_s, \qquad Y_0 = y_0, \qquad (4)$$

where W_s is a Brownian motion and the drift coefficient $b : \mathbb{R} \to \mathbb{R}$ and the diffusion coefficient $\sigma : \mathbb{R} \to \mathbb{R}$ are such that there exists a unique weak solution to (4) (for background on diffusions see, for example, Kloeden and Platen (1995), in particular Chapter 4 for stability conditions). We consider only univariate and time-homogeneous processes.

Suppose that the data

$$Y_{\text{obs}} = (y_0, y_1, \ldots, y_n)$$

comes from Y observed at times $0 = t_0, t_1, \ldots, t_{n-1}, t_n = T$.

The interest is in the parameter θ and we work in the Bayesian setting, assuming a prior $\pi(\theta)$ on θ. Hence, if we denote the transition densities of Y as

$$p(\theta, y_{i-1}, y_i) := \mathbb{P}_\theta(Y_{t_i} \in dy_i | Y_{t_{i-1}} = y_{i-1})/dy_i, \qquad (5)$$

the posterior distribution of θ given data Y_{obs} becomes

$$\pi(\theta|Y_{\text{obs}}) \propto \pi(\theta) \prod_{i=1}^{n} p(\theta, y_{i-1}, y_i). \qquad (6)$$

Unfortunately the diffusion transition density in (5) is not available in closed form except for the constant coefficients case, and a few other very special diffusion families. Consequently, standard inference about θ relies on approximations, such as Euler approximation of the diffusion dynamics, or Monte Carlo approximations of transition densities. We refer to Beskos et al. (2006b), Sermaidis et al. (2013), Gonçalves, Łatuszyński and Roberts (2017a) for discussions of the approximation based inference for diffusions, and in particular the difficult to quantify bias that it introduces, and how the computational cost scales as the desired inferential error decreases.

An alternative to the approximation based inference is the exact inference for diffusions introduced in Beskos et al. (2006b). The approach avoids any discretisation error and allows to design MCMC algorithms that target the exact posterior (6) on an extended state space via introducing an auxiliary variable Y_{mis} that represents the missing continuous paths of Y between discrete observations Y_{obs}. The

first approximation of how such an MCMC algorithm would be designed, is to conceptualise a Gibbs sampling alternating the following two steps:

$$Y_{\mathrm{mis}} \sim \mathcal{L}(Y_{\mathrm{mis}}|Y_{\mathrm{obs}}, \theta); \tag{7}$$

$$\theta \sim \mathcal{L}(\theta|Y_{\mathrm{obs}}, Y_{\mathrm{mis}}). \tag{8}$$

The practical execution of the above Gibbs algorithm, in particular of step (7), is based on exact sampling of diffusion bridges developed in Beskos, Papaspiliopoulos and Roberts (2006a, 2008), termed Exact Algorithm (EA). However, for EA to be applicable, we need to work a bit more. Exact Algorithm is a rejection sampler on the diffusion path space. It proposes a diffusion bridge between two data points from a driftless diffusion (i.e., one with $b \equiv 0$) and accepts or rejects this proposal based on the Radon–Nikodym derivative given by the Girsanov theorem. In (4), the diffusion coefficient σ depends on θ and this presents a problem as there is a perfect correlation between the diffusion path (precisely speaking, its quadratic variation) and the diffusion coefficient, so the Gibbs sampler above would not mix at all. Second, the driftless diffusion bridge proposals for Y_{mis} in the EA need to be simulable, which in practice means that the diffusion coefficient in (4) can not depend on Y_s.

Hence, in order to perform MCMC inference, the diffusion (4) is transformed into a stochastic differential equation with unit diffusion coefficient using Lamperti transformation. This allows us to obtain a parameter-free dominating measure to write the likelihood function of a complete diffusion path, and in particular to design diffusion Bridge proposals based on the standard Brownian motion. The Lamperti transform

$$X_s = \eta(Y_s; \theta) = \int_v^{Y_s} \frac{1}{\sigma(u; \theta)} \, du, \tag{9}$$

where v is some arbitrary element of the state space of Y, implies that X is a diffusion solving the SDE

$$dX_s = \alpha(X_s; \theta) \, ds + dW_s, \qquad X_0 = x_0(\theta) = \eta(y_0, \theta), \tag{10}$$

where

$$\alpha(u; \theta) = \frac{b(\eta^{-1}(u; \theta); \theta)}{\sigma(\eta^{-1}(u; \theta); \theta)} - \frac{\sigma'(\eta^{-1}(u; \theta); \theta)}{2}. \tag{11}$$

Note that X_s is a function of θ, in particular the observed data becomes

$$\big(x_0(\theta), \ldots, x_n(\theta)\big) = \big(\eta(y_0; \theta), \ldots, \eta(y_n; \theta)\big).$$

The second transformation, now only for the bridges of X (between the observations), is given by

$$\dot{X}_s = \varphi_\theta^{-1}(X_s)$$
$$:= X_s - \left(1 - \frac{s - t_{i-1}}{t_i - t_{i-1}}\right) x_{i-1}(\theta) - \left(\frac{s - t_{i-1}}{t_i - t_{i-1}}\right) x_i(\theta), \tag{12}$$

for $s \in (t_{i-1}, t_i)$. This transformation guarantees that the measure of each bridge $\dot{X}_s$ is dominated by the measure of a standard Brownian bridge of same time length. We denote the transformation in (12) as φ_θ^{-1}, so that we have the handy notation $\varphi_\theta(\dot{X}_s)$ for its inverse.

By $\dot{X}_{\mathrm{mis}}$ denote the bridges of $\dot{X}$. Defining $Y_{\mathrm{com}} = \{Y_{\mathrm{obs}}, \dot{X}_{\mathrm{mis}}\}$, where $Y_{\mathrm{obs}} = (y_0, y_1, \ldots, y_n)$ and X_{mis} are all the bridges of $\dot{X}$, Lemma 2 from Beskos et al. (2006b) gives the likelihood $L(\theta; Y_{\mathrm{com}})$ of a complete diffusion path in $[0, T]$ and the joint posterior density of $(\theta, X_{\mathrm{mis}})$ satisfies $\pi(\theta, X_{\mathrm{mis}}|Y_{\mathrm{obs}}) \propto L(\theta; Y_{\mathrm{com}})\pi(\theta)$, which gives us the full conditional densities for the Gibbs sampler, as detailed in Sections 4.2 and 4.3.

Exact methodologies for inference in diffusion processes have been firstly proposed in Beskos et al. (2006b) and are based on the Exact Algorithm (EA), which simulates a class of diffusion processes exactly via rejection sampling. However, the feasibility of all those methodologies rely on the assumption that a given function of the diffusion components is bounded below (see condition (13) below), while their efficiency depends on the tightness of this bound. Although broad, the class of processes that satisfy this assumption exclude some appealing processes, such as the Cox–Ingersoll–Ross model and Wright–Fisher family of diffusions (Jenkins and Spanó (2016)). The methodology we present in this section is general enough not to require such assumption and, therefore, can be applied to models such as the ones just mentioned. We present an example with a model from the Wright–Fisher family.

As mentioned, the existing exact inference methodologies rely heavily on the assumption that a certain function is bounded below. The function in question is $(\alpha^2 + \alpha')(u; \theta)$, where α is defined in (11), and the derivative is in the space variable u. In case of one dimensional real valued diffusions, the precise condition reads

$$\inf_{u \in \mathbb{R}} \{(\alpha^2 + \alpha')(u; \theta)\} \geq a(\theta) > -\infty. \tag{13}$$

Hence the function is required to be uniformly bounded below in the state space of X, for all θ in the parametric space. The methodology presented here, however, does not require this boundedness assumption, as we make it clear further ahead in this section.

4.2 Sampling the missing paths

The transformed missing bridges X_{mis} are sampled via Barker's with standard Brownian bridge proposals. Given that the bridges are conditionally independent, due to the Markov property, the Barker's step for each bridge may be performed in parallel. Our description shall focus on the update step of a single interval: (t_{i-1}, t_i).

The Barker's acceptance probability α_X of a proposed bridge $\dot{B}$ given a current bridge $\dot{X}$ in (t_{i-1}, t_i) is obtained using the measure of a standard Brownian bridge as the dominating measure and is given by

$$\alpha_X = \alpha_X(\dot{X}, \dot{B}) = \frac{G(\varphi_\theta(\dot{B}); \theta)}{G(\varphi_\theta(\dot{B}); \theta) + G(\varphi_\theta(\dot{X}); \theta)},$$

where G is derived from Girsanov's formula (see, for example, Sermaidis et al. (2013) in a similar context): for an arbitrary path Y

$$G(Y; \theta) = \exp\left\{-\int_{t_{i-1}}^{t_i} \left(\frac{\alpha^2 + \alpha'}{2}\right)(Y_s; \theta)\, ds\right\}.$$

Now we intend to follow the strategy of conditioning by an auxiliary variable adopted in the simple example of Section 3. However, to do this we require an upper bound on G. For some diffusion models, the functional form of $\alpha^2 + \alpha'$ is bounded below which leads to an upper bound for G. However for many models (including the Wright–Fisher model we shall go on to consider) we will need to have additional information about the sample path Y to provide the necessary bounds. To this end, we adopt the *layered Bownian bridge* construction of Beskos et al. (2006b), Beskos, Papaspiliopoulos and Roberts (2008). Giving a detailed description of this construction is beyond the scope of this paper though the complete details can be found in Beskos, Papaspiliopoulos and Roberts (2008), Gonçalves, Łatuszyński and Roberts (2017a). The important feature of this construction for our purposes is that the layer of a path, in particular of the proposal path $\dot{B}$, can be directly simulated, and leads to upper and lower bounds on its potential trajectories that are of the form $L < \inf_{s \in (t_{i-1}, t_i)} \dot{B}_s < \sup_{s \in (t_{i-1}, t_i)} \dot{B}_s < U$ for some $L, U \in \mathbb{R}$. This in turn allows us to produce a local lower bound on of $\alpha^2 + \alpha'$ on the compact set $[L, U]$ which apply for any path consistent with that layer. For trajectory X, we shall call this lower bound $2a_i(X; s, \theta)$, so that

$$\frac{(\alpha^2 + \alpha')(\varphi_\theta(\dot{X}_s))}{2} \geq a_i(X; s, \theta) \tag{14}$$

for all paths X consistent with the layer of X. Note that a_i may not depend upon s as we are here only considering $s \in [t_{i-1}, t_i]$ although in the next section we shall need to consider dependence on s as X will have different layers in different intervals. Furthermore, tighter bounds may be obtained if the layers are obtained for a standard bridge and then transformed to the original one, making the bounds a linear function of s. Thus we shall write

$$\alpha_X = \frac{s_B p_B}{s_B p_B + s_X p_X}, \qquad \text{where} \tag{15}$$

$$s_X = \exp\{-\mathcal{I}_{a_i}(\theta)\}, \tag{16}$$

$$p_X = \exp\left\{-\int_{t_{i-1}}^{t_i} \left(\frac{\alpha^2 + \alpha'}{2}\right)(\varphi_\theta(\dot{X}_s); \theta) - a_i(X; s, \theta)\, ds\right\},$$

$$\mathcal{I}_{a_i}(\theta) = \int_{t_{i-1}}^{t_i} a_i(X; s, \theta)\, ds, \qquad \text{moreover,} \tag{17}$$

$$a_i(X; s, \theta) \le \left(\frac{\alpha^2 + \alpha'}{2}\right)(\varphi_\theta(\dot{X}_s); \theta), \qquad s \in [t_{i-1}, t_i]. \tag{18}$$

Consequently, s_B and s_X are known positive numbers and p_B and p_X are unknown probabilities. However, coins with probabilities p_B and p_X can be simulated using an algorithm called the Poisson coin. We refer to Beskos, Papaspiliopoulos and Roberts (2006a, 2008), Gonçalves, Łatuszyński and Roberts (2017a) for a detailed construction and use of the Poisson coin, the brief description is as follows.

Note that by (14) the integrand in (17) is strictly positive and the layered Brownian bridge construction will also yield a bound $r_i(X, \theta) \in \mathbb{R}_+$, valid for any path consistent with the simulated layer, and such that

$$r_i(X, \theta) \ge \sup_{s \in (t_{i-1}, t_i)} \left\{ \left(\frac{\alpha^2 + \alpha'}{2}\right)(\varphi_\theta(\dot{X}_s); \theta) - a_i(X; s, \theta) \right\}. \tag{19}$$

Now let Φ be a Poisson process with intensity $r_i(X, \theta)$ on $[t_{i-1}, t_i] \times [0, 1]$. Its realisation is a collection of points $\{\phi_k, \chi_k\}_{k=1}^{\kappa}$ on $[t_{i-1}, t_i] \times [0, 1]$, where $\kappa \sim \text{Poiss}[(t_i - t_{i-1}) r_i(X, \theta)]$. If N is the number of points below the graph

$$s \longrightarrow \frac{(\frac{\alpha^2 + \alpha'}{2})(\varphi_\theta(\dot{X}_s); \theta) - a_i(X; s, \theta)}{r_i(X, \theta)},$$

then

$$\mathbb{P}[N = 0 | X] = p_X.$$

Sampling the Poisson process Φ is standard and verifying if its points are below or above the graph requires revealing $\dot{X}_s$ at a finite collection of timepoints $\phi_1, \ldots, \phi_\kappa$, which in the case of Brownian bridge proposals is also routine.

Thus all the steps of the 2-coin algorithm can be performed in order to sample an event of probability α_X and accept or reject the proposed Brownian bridge $\dot{B}$ in the missing path update step of the Gibbs sampler.

4.3 Sampling the parameters

The parameter vector θ is also sampled by a Barker's step. The proposal distribution is a symmetric random walk, but not necessarily Gaussian. The Barker's acceptance probability α_θ for a proposal θ^*, given a current value θ, is obtained using the dominating measure from the prior density $\pi(\theta)$ and is given by

$$\alpha_\theta = \frac{s_{\theta^*} p_{\theta^*}}{s_{\theta^*} p_{\theta^*} + s_\theta p_\theta}, \qquad \text{where}$$

$$s_\theta = \exp\left\{ A(x_n(\theta); \theta) - A(x_0(\theta); \theta) - \sum_{i=1}^{n} \mathcal{I}_{a_i}(\theta) \right\} \pi(\theta),$$

$$p_\theta = \exp\left\{ -\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} \left(\frac{\alpha^2 + \alpha'}{2}\right)(\varphi_\theta(\dot{X}_s); \theta) - a_i(X; s, \theta)\, ds \right\},$$

$$A(u; \theta) = \int_0^u \alpha(y; \theta)\, dy.$$

Once again, the acceptance decision of probability α_θ can be performed using the 2-coin algorithm and the Poisson coin in a manner analogous to that described in Section 4.2. The efficiency of the proposed methodology relies heavily on the probability of the two second coins which, in turn, relies on the lower bounds $a_i(X; s, \theta)$. A detailed description of how to obtain efficient lower bounds $a_i(X; s, \theta)$ can be found in Gonçalves, Łatuszyński and Roberts (2017a).

Finally, note that our MCMC steps do not require function $(\alpha^2 + \alpha')$ to be bounded below. The form of the Barker's acceptance probability and the dynamics of the 2-coin algorithm requires only that bounds for the diffusion path are obtained. More specifically, the bounds for the Brownian bridge construction are used to obtain a lower bound for $(\alpha^2 + \alpha')$ which allows us to write the Barker's acceptance probability in the 2-coin algorithm form.

4.4 Example

The Wright–Fisher family of diffusions

$$dY_s = \beta(Y_s)\, ds + \sqrt{Y_s(1 - Y_s)}\, dW_s$$

is widely used in statistical applications, especially in genetics, see Schraiber, Griffiths and Evans (2013) and references therein. It is an example for which after applying the lamperti transformation the function $(\alpha^2 + \alpha')$ of the resulting drift α is not bounded below. Jenkins and Spanó (2016) propose algorithms to perform exact simulation of processes in that family but, although this could potentially be used to developed exact inference methodology, the authors do not pursue this direction in their paper.

We apply the Barker's methodology presented above to perform exact inference about drift parameters for the neutral Wright–Fisher diffusion with mutation which admits the following parametric SDE:
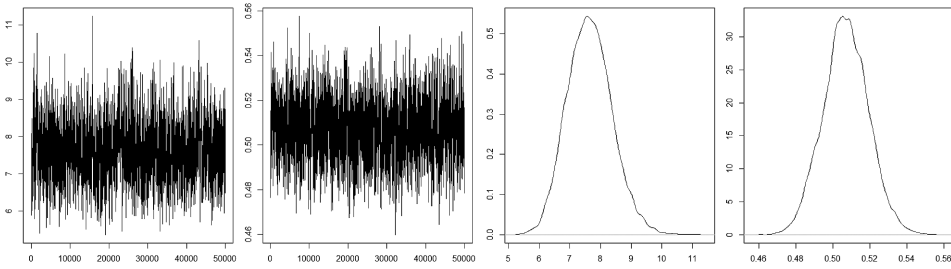
$$dY_s = \frac{1}{2}(\theta_1(1 - Y_s) - \theta_2 Y_s)\, ds + \sqrt{Y_s(1 - Y_s)}\, dW_s, \tag{20}$$

$$Y_0 = y_0, \theta_1, \theta_2 > 0.$$

For inference purposes, we consider the following reparametrisation:

$$\gamma_1 = \theta_1 + \theta_2 \quad \text{and} \quad \gamma_2 = \frac{\theta_1}{\theta_1 + \theta_2}.$$

The two new parameters γ_2 and γ_1 represent the process' reversible mean and the drift force towards it, respectively, and are expected to have low posterior correlation. This is useful since it makes an uncorrelated random walk proposal for

Figure 2 *Posterior distribution of the parameter set: trace plots and marginal posterior densities.*

Table 1 *Posterior statistics for the parameters*

	Mean	s.d.	95% C.I.
γ_1	7.649	0.729	(6.502, 8.895)
γ_2	0.507	0.012	(0.486, 0.527)

(γ_1, γ_2) a reasonable choice. After this reparametrisation, the Lamperti transform leads to a diffusion with unit diffusion coefficient and drift of the form

$$\alpha(u; \theta) = \frac{1}{2 \sin(u)} (\gamma_1 (2\gamma_2 - 1) + (\gamma_1 - 1) \cos(u)), \qquad u \in (0, \pi).$$

One can check that $(\alpha^2 + \alpha')$ is not uniformly bounded below in the state space for some region of the parametric space, however it can be bounded conditionally once upper and lower bounds for the diffusion path are obtained through the layered Brownian bridge construction discussed above.

We simulate 201 equally spaced observations in $[0, 200]$ for $\gamma_1 = 8$ and $\gamma_2 = 0.5$. Uniform independent priors on the positive real line are adopted for γ_1 and γ_2. The chain runs for 50 thousand iterations with two consecutive updates for $\dot{X}$ at each iteration to improve the mixing of the chain. The proposal distribution for (γ_1, γ_2) is a uniform random walk for each coordinate—$U(\gamma_1 \pm 0.65)$ and $U(\gamma_2 \pm 0.01)$. The acceptance rate of the Barker's step for the parameter vector was 0.357. The estimated posterior correlation of γ_1 and γ_2 was -0.005. Results are presented in Figure 2 and Table 1. A burn-in of 2 thousand iterations is used to compute the posterior estimates.

Acknowledgments

References

Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics* **37**, 697–725. MR2502648

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

Asmussen, S., Glynn, P. and Thorisson, H. (1992). Stationarity detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation* **2**, 130–157.

Barker, A. A. (1965). Monte Carlo calculations of the radial distribution functions for a protonelectron plasma. *Australian Journal of Physics* **18**, 119–133.

Beaumont, M. A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics* **164**, 1139–1160.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006a). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12**, 1077–1098.

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2008). A new factorisation of diffusion measure and sample path reconstruction. *Methodology and Computing in Applied Probability* **10**, 85–104.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O. and Fearnhead, P. (2006b). Exact and computationally efficient likelihood-based inference for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society, Series B* **68**, 333–382.

Flegal, J. M. and Herbei, R. (2012). Exact sampling for intractable probability distributions via a Bernoulli factory. *Electronic Journal of Statistics* **6**, 10–37.

Gonçalves, F. B., Łatuszyński, K. G. and Roberts, G. O. (2017a). Exact Monte Carlo likelihood-based inference for jump-diffusion processes. Available at arXiv:1707.00332.

Gonçalves, F. B., Roberts, G. O. and Łatuszyński, K. G. (2017b). Exact Bayesian inference for diffusion driven Cox processes. In preparation.

Herbei, R. and Berliner, L. M. (2014). Estimating ocean circulation: An MCMC approach with approximated likelihoods via the Bernoulli factory. *Journal of the American Statistical Association* **109**, 944–954.

Huber, M. (2015). Optimal linear Bernoulli factories for small mean problems. Available at arXiv:1507.00843.

Huber, M. (2016). Nearly optimal Bernoulli factories for linear functions. *Combinatorics, Probability & Computing* **25**, 577–591.

Jacob, P. E. and Thiery, A. H. (2015). On nonnegative unbiased estimators. *The Annals of Statistics* **43**, 769–784.

Jenkins, P. A. and Spanó, D. (2016). Exact simulation of the Wright–Fisher diffusion. *The Annals of Applied Probability* **27**, 1478–1509. MR3678477

Keane, M. and O'Brien, G. (1994). A Bernoulli factory. *ACM Transactions on Modeling and Computer Simulation* **4**, 213–219.

Kloeden, P. and Platen, E. (1995). *Numerical Solution of Stochastic Differential Equations*. New York: Springer.

Łatuszyński, K., Kosmidis, I., Papaspiliopoulos, O. and Roberts, G. (2011). Simulating events of unknown probabilities via reverse time martingales. *Random Structures & Algorithms* **38**, 441–452.

Łatuszyński, K., Palczewski, J. and Roberts, G. (2017). Exact inference for a Markov switching diffusion model with discretely observed data. In preparation.

Łatuszyński, K. and Roberts, G. O. (2013). CLTs and asymptotic variance of time-sampled Markov chains. *Methodology and Computing in Applied Probability* **15**, 237–247.

Mira, A. (2001). Ordering and improving the performance of Monte Carlo Markov chains. *Statistical Science* **16**, 340–350. MR1888449

Mira, A. and Geyer, C. (1999). Ordering Monte Carlo Markov chains. Technical report, School of Statistics, Univ. Minnesota.

Nacu, Ş. and Peres, Y. (2005). Fast simulation of new coins from old. *The Annals of Applied Probability* **15**, 93–115.

Papaspiliopoulos, O. and Roberts, G. O. (2008). Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. *Biometrika* **95**, 169–186.

Peskun, P. H. (1973). Optimum Monte Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.

Roberts, G. and Rosenthal, J. (1997). Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability* **2**, 13–25.

Schraiber, J. G., Griffiths, R. C. and Evans, S. N. (2013). Analysis and rejection sampling of Wright–Fisher diffusion bridges. *Theoretical Population Biology* **89**, 64–74.

Sermaidis, G., Papaspiliopoulos, O., Roberts, G. O., Beskos, A. and Fearnhead, P. (2013). Markov chain Monte Carlo for exact inference for diffusions. *Scandinavian Journal of Statistics* **40**, 294–321. MR3066416

Von Neumann, J. (1951). Various techniques used in connection with random digits. In *Monte Carlo Method, Vol. 12*. National Bureau of Standards.

F. B. Gonçalves
Departamento de Estatística
Universidade Federal de Minas Gerais
Av. Antônio Carlos
6627—DEST/ICEx/UFMG—Belo Horizonte
Minas Gerais, 31270-901
Brazil
E-mail: fbgoncalves@est.ufmg.br

K. Łatuszyński
Department of Statistics
University of Warwick
Coventry CV4 7AL
UK
and
Alan Turing Institute
British Library
96 Euston Rd
London NW1 2DB
UK

G. O. Roberts
Department of Statistics
University of Warwick
Coventry CV4 7AL
UK