The Annals of Statistics
2015, Vol. 43, No. 2, 769–784
DOI: 10.1214/15-AOS1311

ON NONNEGATIVE UNBIASED ESTIMATORS

BY PIERRE E. JACOB[1,2] AND ALEXANDRE H. THIERY[2]

University of Oxford and National University of Singapore

We study the existence of algorithms generating almost surely nonnegative unbiased estimators. We show that given a nonconstant real-valued function f and a sequence of unbiased estimators of $\lambda \in \mathbb{R}$, there is no algorithm yielding almost surely nonnegative unbiased estimators of $f(\lambda) \in \mathbb{R}^+$. The study is motivated by pseudo-marginal Monte Carlo algorithms that rely on such nonnegative unbiased estimators. These methods allow "exact inference" in intractable models, in the sense that integrals with respect to a target distribution can be estimated without any systematic error, even though the associated probability density function cannot be evaluated pointwise. We discuss the consequences of our results on the applicability of pseudo-marginal algorithms and thus on the possibility of exact inference in intractable models. We illustrate our study with particular choices of functions f corresponding to known challenges in statistics, such as exact simulation of diffusions, inference in large datasets and doubly intractable distributions.

1. Introduction.

1.1. *Exact inference through unbiased estimators.* Consider the problem of estimating the integral of a function φ with respect to a probability distribution with density π. A successful Markov chain Monte Carlo or sequential Monte Carlo method allows us to estimate integrals with respect to π in such a way that the error can be reduced down to zero by producing more samples. We call these methods "exact" since there is no systematic error in the estimation, even though the sampling error can be large for a given computational budget. Using the Metropolis–Hastings algorithm, exact inference is possible when the target probability density function π can be evaluated pointwise up to a multiplicative constant.

The possibility of performing exact inference without relying on evaluations of the target probability density function is an important open question. A class of exact methods, called pseudo-marginal Metropolis–Hastings, has been proposed in Andrieu and Roberts (2009), generalizing and validating methods developed in population genetics [Beaumont (2003)] and lattice quantum chromodynamics

Received May 2014; revised October 2014.
[1]Supported by EPSRC through Grant EP/K009362/1.
[2]Supported by a Singapore Ministry of Education grant.
MSC2010 subject classifications. 65C50, 65C60, 68W20.
Key words and phrases. Unbiased estimator, Poisson estimator, Monte Carlo methods, sign problem, Bernoulli factory.

[Kennedy and Kuti (1985)]. Pseudo-marginal methods rely on nonnegative unbiased estimators of density evaluations $\pi(x)$ instead of the evaluations themselves. In a related manner, Del Moral, Doucet and Jasra (2007), Fearnhead, Papaspiliopoulos and Roberts (2008), Fearnhead et al. (2010), Liu and Chen (1998), Tran et al. (2013) show that sequential Monte Carlo methods remain exact when the importance weights are replaced by nonnegative unbiased estimators thereof.

The applicability of exact methods has thus been considerably extended since estimating $\pi(x)$ is generally easier than evaluating it. For instance, in the common case where the cost of evaluating the likelihood function grows at least linearly with the size of the dataset, pointwise posterior density evaluations become prohibitive for large datasets but can potentially be estimated using subsampling [Kleiner et al. (2014), Welling and Teh (2011)]. In state space models, the likelihood involves an intractable integral over a latent stochastic process but can be estimated using particle filters [Andrieu, Doucet and Holenstein (2010)]. In other settings, the likelihood cannot be evaluated because it involves an intractable normalizing constant, such as in "doubly intractable" models commonly found in spatial statistics and graphical models [Everitt (2012), Girolami et al. (2013), Møller et al. (2006)]. Even for simple models and small datasets, the use of reference priors for an objective Bayesian analysis leads to posterior probability density functions that cannot be evaluated pointwise [Berger, Bernardo and Sun (2009)] for they involve limits or infinite sums. In each case, exact inference can still be achieved through a pseudo-marginal approach, provided that an appropriate nonnegative unbiased estimator $\widehat{\pi}(x)$ is available.

Generic techniques to obtain unbiased estimators from biased ones, referred to as "debiasing techniques," have been developed independently in various fields and recently reviewed and generalized in McLeish (2011), Rhee and Glynn (2012, 2013). The combination of debiasing techniques and pseudo-marginal methods provides a promising roadmap to perform exact inference in a very general setting. Unfortunately unbiased estimators $\widehat{\pi}(x)$, as produced by current debiasing techniques, can take negative values with positive probability, even if their expectations $\pi(x)$ are known to be nonnegative. These negative values prevent the direct use of unbiased estimators within a pseudo-marginal Markov chain algorithm. Likewise, standard sequential Monte Carlo methods cannot be directly implemented when negative values can be encountered.

One might want to avoid the sign problem completely by using unbiased estimators that only take nonnegative values. In other words, one might hope to find a debiasing technique which satisfies a sign constraint. We propose to study the design of such algorithms. In Section 1.2 we recall the main ideas behind debiasing techniques and highlight the connection with the Bernoulli factory [Keane and O'Brien (1994)]. In Section 1.3 we describe applications in statistics. In Section 2 we present a result stating the nonexistence of generic schemes to obtain nonnegative unbiased estimators. In Section 3 we discuss their existence under additional conditions, which in practice require additional model-specific information. The results and further research venues are discussed in Section 4.

1.2. *Designing unbiased estimators.* Our results are connected to the literature on debiasing techniques and Bernoulli factories. In computational physics, Kuti (1982) uses a method to unbiasedly estimate some elements of the inverse of a matrix without fully inverting it, while Wagner (1987) proposes unbiased estimators of functional integrals; both methods are inspired by an unpublished scheme of J. von Neumann and S. M. Ulam. A similar idea has been proposed by Rychlik (1990) for estimating the derivative of a regression function and by Rychlik (1995) for kernel density estimation. More recently McLeish (2011) and then Rhee and Glynn (2012, 2013) have proposed a general scheme to remove the bias of a sequence of consistent estimators $(S_n)_{n\geq 0}$ of a quantity $\lambda \in \mathbb{R}$, satisfying

$$\lim_{n\to\infty} \mathbb{E}(S_n) = \mathbb{E}(S) = \lambda. \tag{1.1}$$

In equation (1.1), the quantity S can either be thought of as a random variable that is impossible to generate in finite time and S_n as an approximation of S, or simply as the desired, and generally unknown value $S = \lambda$. Suppose that one can sample from S_n for each $n \geq 0$. Let N be an integer-valued random variable that is independent of the sequence $(S_n)_{n\geq 0}$ and that can take arbitrary large values. Under mild assumptions, with the convention $S_{-1} = 0$, the weights $w_n = 1/\mathbb{P}(N \geq n)$ are such that the random sum

$$Y = \sum_{n=0}^{N} w_n \times (S_n - S_{n-1}) \tag{1.2}$$

is an unbiased estimator of λ. The following result gives a condition for its second moment to be finite.

THEOREM 1.1 [Theorem 1 of Rhee and Glynn (2013)]. *Introduce a random variable S with $\mathbb{E}(S) = \lambda \in \mathbb{R}$. Let $(S_n)_{n\geq 0}$ be a sequence of random variables, let N be an integer valued random variable that can take arbitrarily large values and set $w_n = 1/\mathbb{P}(N \geq n)$. Under the condition*

$$\sum_{n=1}^{\infty} w_n \times \mathbb{E}\big(|S - S_{n-1}|^2\big) < \infty, \tag{1.3}$$

the random variable $Y = \sum_{n=0}^{N} w_n \times (S_n - S_{n-1})$, with the convention $S_{-1} = 0$, is well defined, has expectation $\mathbb{E}(Y) = \mathbb{E}(S) = \lambda$ and a finite second moment

$$\mathbb{E}(Y^2) = \sum_{n=0}^{\infty} w_n \times \big(\mathbb{E}\big(|S - S_{n-1}|^2\big) - \mathbb{E}\big(|S - S_n|^2\big)\big) < \infty.$$

The "debiased" estimator Y also generalizes the random truncation approach discussed in Girolami et al. (2013), Papaspiliopoulos (2011) and references therein. The random variable N could be replaced by a stopping time. Since the

random sum in equation (1.2) only involves an almost surely finite number of terms, the estimator Y is straightforward to simulate.

In the case where the quantity of interest λ is nonnegative, the random sum in equation (1.2) can still take negative values, even if the original estimators $(S_n)_{n\geq 0}$ were all almost surely nonnegative; this is because each increment $(S_n - S_{n-1})$ can potentially be negative. An important exception occurs when the estimators $(S_n)_{n\geq 0}$ are ordered, that is, $S_n \geq S_{n-1}$ almost surely. With exact inference in mind, one can wonder about the existence of other debiasing techniques which, unlike Y of equation (1.2), would only yield nonnegative values. Section 2 will introduce a framework to study that question.

Our framework will also be related to Bernoulli factories, which have been introduced in the seminal article of Keane and O'Brien (1994). Given a subset $\mathcal{P} \subset [0, 1]$ and a function $f : \mathcal{P} \to [0, 1]$, a Bernoulli factory generates Bernoulli random variables with success probability $f(p)$ given as input an independent sequence of Bernoulli random variables with success probability $p \in \mathcal{P}$; of course the algorithm does not have access to the value p. The existence of such an algorithm depends on the subset $\mathcal{P}$ and on the function f considered. For instance, there does not exist an algorithm for $f : p \mapsto 2p$ and $\mathcal{P} = [0, 1/2]$; maybe surprisingly, there does exist an algorithm for the same function f and the set $\mathcal{P} = [0, 1/2 - \varepsilon]$ for any $\varepsilon > 0$. It will become apparent in Section 3.2 that the construction of nonnegative unbiased estimators shares many similarities with the Bernoulli factory.

To summarize, debiasing techniques allow us to construct unbiased estimators of generic quantities but do not ensure that the resulting estimates are nonnegative. On the other hand Bernoulli factories always produce nonnegative variables, but require Bernoulli variables as input. In general we are interested in the existence of algorithms producing nonnegative unbiased estimators of $f(\mathbb{E}[X])$ for some function $f : \mathbb{R} \to \mathbb{R}^+$ and real-valued random variables X.

1.3. *Applications.* Some specific choices of function f are of special interest in applied probability and statistics, especially the exponential $f : x \mapsto \exp(x)$ and the inverse $f : x \mapsto 1/x$.

The exponential case appears whenever log-likelihood evaluations can be unbiasedly estimated. An algorithm generating unbiased estimates of $\exp(\lambda)$ from a stream of unbiased estimates of λ is referred to as a *Poisson estimator* in the literature on perfect simulation and inference for diffusion processes [Beskos, Papaspiliopoulos and Roberts (2006), Beskos and Roberts (2005), Beskos et al. (2006), Fearnhead, Papaspiliopoulos and Roberts (2008), Jourdain and Sbai (2007), Olsson and Ströjby (2011), Sermaidis et al. (2015)], and the first occurrence might be in Bhanot and Kennedy (1985). On a finite interval $[0, T]$, the probability distribution $\mathbb{Q}$ on the space of continuous functions $\mathcal{C}([0, T], \mathbb{R})$ generated by a scalar diffusion processes with unit volatility coefficient $dX_t = \mu(X_t)\,dt + dW_t$ has, under mild regularity assumptions on the drift function

$\mu : \mathbb{R} \to \mathbb{R}$, a Radon–Nikodym derivative with respect to the standard Wiener measure $\mathbb{W}$ that can be expressed as

$$\frac{d\mathbb{Q}}{d\mathbb{W}}\big((x_t)_{t=0}^T\big) = \exp\bigg(\int_{t=0}^T \Phi(x_t)\,dt\bigg)$$

for an explicit function $\Phi : \mathbb{R} \to \mathbb{R}$ given by Girsanov's theorem. As described in Beskos et al. (2006), unbiased estimates of the integral $\int_{t=0}^T \Phi(x_t)\,dt$ can be obtained by standard importance sampling. The existence of a Poisson estimator allows us to transform these samples into an unbiased estimate of $(d\mathbb{Q}/d\mathbb{W})((x_t)_{t=0}^T)$, which can then be used for exact inference.

The exponential case also appears in the context of inference for large datasets, where the posterior probability density function π is expensive to evaluate pointwise. Indeed the log-likelihood $\ell(\theta) = \sum_{i=1}^n \log f(y_i \mid \theta)$ of $n \gg 1$ independent observations $(y_i)_{i=1}^n$ can be unbiasedly estimated at reduced cost by using a random subsample of only $m \ll n$ observations. For instance, given any $m \geq 1$, the quantity $\widehat{\ell}(\theta) = (n/m)\sum_{i=1}^m \log f(y_{\sigma_i} \mid \theta)$, where $(\sigma_i)_{i=1}^m$ are drawn uniformly in $\{1, \ldots, n\}$, is an unbiased estimator of $\ell(\theta)$.

The choice $f : x \mapsto 1/x$ appears in the context of doubly intractable models [Girolami et al. (2013), Walker (2011)] where the observations are assumed to follow a distribution with density

$$f(y \mid \theta) = \frac{g(y,\theta)}{\int g(s,\theta)\,ds}$$

for a function $(y,\theta) \mapsto g(y,\theta)$ that can be evaluated pointwise. The denominator $Z(\theta) = \int g(s,\theta)\,ds$ is generally intractable, which prevents the use of the standard Metropolis–Hastings algorithm to obtain posterior estimates. Nevertheless $Z(\theta)$ can be unbiasedly estimated by standard importance sampling. Would a nonnegative estimator of $1/Z(\theta)$ be available, a pseudo-marginal approach could be implemented.

2. Existence of nonnegative unbiased estimation schemes.

2.1. *Algorithms and factories.* For any nonempty measurable space $\mathcal{X} \subset \mathbb{R}$, let $\mathcal{M}_1(\mathcal{X})$ be the set of probability distributions on $\mathcal{X}$ with finite first moment and $\operatorname{conv}(\mathcal{X})$ the smallest interval containing $\mathcal{X}$. For $\mu \in \mathcal{M}_1(\mathcal{X})$ we use the notation $m_1(\mu) = \int_{\mathcal{X}} x\mu(dx)$ for the mean of μ; indeed, $m_1(\mu) \in \operatorname{conv}(\mathcal{X})$ for any $\mu \in \mathcal{M}_1(\mathcal{X})$. The distribution of the random variable X is denoted by $\mathcal{D}(X)$. Let $L^2(\mathcal{X})$ be the space of square integrable random variables on $\mathcal{X}$. The indicator function of a set A is denoted by 1_A, and 1_x for some $x \in \mathbb{R}$ denotes the Dirac delta function centered at x. An unbiased estimator of a quantity $\lambda \in \mathbb{R}$ is called a U-estimator of λ, or a U^+-estimator in the case where it is almost surely nonnegative.

For a function $f : \operatorname{conv}(\mathcal{X}) \to \mathbb{R}^+$, we propose to study the existence of f-factories, defined as devices taking as input U-estimators of $\lambda \in \operatorname{conv}(\mathcal{X})$ with

support on $\mathcal{X}$, and producing U^+-estimators of $f(\lambda)$. Borrowing ideas from Keane and O'Brien (1994), we first define rigorously a class of algorithms that we will consider practical.

DEFINITION 2.1. Let $\mathcal{X}$ be a subset of $\mathbb{R}$. An $\mathcal{X}$-algorithm $\mathcal{A}$ is a pair (T, φ) where $T = (T_n)_{n\geq 1}$ is a sequence of functions $T_n : (0,1) \times \mathcal{X}^n \to \{0,1\}$, and $\varphi = (\varphi_n)_{n\geq 1}$ is a sequence of functions $\varphi_n : (0,1) \times \mathcal{X}^n \to \mathbb{R}^+$.

An $\mathcal{X}$-algorithm $\mathcal{A} \equiv (T, \varphi)$ takes an infinite sequence $x = (x_n)_{n\geq 1} \in \mathcal{X}^\infty$ and an auxiliary variable $u \in (0,1)$ as input and produces as output

$$\mathcal{A}(u, x) = \varphi_\tau(u, x_1, \ldots, x_\tau)$$

with $\tau = \tau(u, x) = \inf\{n \geq 1 : T_n(u, x_1, \ldots, x_n) = 1\}$. We adopt the convention $\mathcal{A}(u, x) = \infty$ when $\{n \geq 1 : T_n(u, x_1, \ldots, x_n) = 1\} = \varnothing$ and say in this case that the algorithm does not terminate. In the applications that we have in mind, the infinite sequence $x = (x_n)_{n\geq 1} \in \mathcal{X}^\infty$ is the realization of an independent sequence of random variables $X = (X_n)_{n\geq 1}$, and the variable $u \in (0,1)$ is the realization of a random variable $U \sim \text{Uniform}(0,1)$ independent of X. In this case, we say that the algorithm almost surely terminates if $\mathbb{P}(\tau < \infty) = 1$. Definition 2.1 translates the fact that a valid algorithm uses a possibly random amount of inputs and that the decision to stop acquiring more inputs only relies on the information contained in the already acquired inputs.

The variable U allows the algorithm to be randomized: on top of the sequence $(X_n)_{n\geq 1}$ it can sample additional random variables. Specifying a single auxiliary variable $U \sim \text{Uniform}(0,1)$ or an infinite independent sequence $(U_n)_{n\geq 1}$ of uniforms is equivalent. Indeed, one can construct an infinite sequence of independent Bernoulli random variables by considering the binary expansion of $U \sim \text{Uniform}(0,1)$, and then partition the expansion into disjoint infinite subsequences to obtain an infinite number of binary representations of independent uniform random variables.

DEFINITION 2.2. Let $\mathcal{X}$ be a subset of $\mathbb{R}$ and $f : \text{conv}(\mathcal{X}) \to \mathbb{R}^+$ a function. An f-factory $\mathcal{A} \equiv (\varphi, T)$ is an $\mathcal{X}$-algorithm such that for any distribution $\pi \in \mathcal{M}_1(\mathcal{X})$, an independent sequence $X = (X_n)_{n\geq 1}$ marginally distributed as π and an auxiliary random variable $U \sim \text{Uniform}(0,1)$ independent of $(X_n)_{n\geq 1}$, the random variable $Y = \mathcal{A}(U, X)$ is a nonnegative unbiased estimator of $f(m_1(\pi))$.

The condition $\mathbb{E}(\mathcal{A}(U, X)) = f(m_1(\pi))$ implies that the algorithm terminates with probability one when fed with the independent sequence $X = (X_n)_{n\geq 1}$ and $U \sim \text{Uniform}(0,1)$. Importantly the definition implies that an f-factory should work for any distribution $\pi \in \mathcal{M}_1(\mathcal{X})$.

2.2. *Nonexistence of general f-factories.* We first consider the general case $\mathcal{X} = \mathbb{R}$, where the unbiased estimators used as input can take any real value.

THEOREM 2.1. *For any nonconstant function $f : \mathbb{R} \to \mathbb{R}^+$, no f-factory exists.*

PROOF. For the sake of contradiction, suppose that there exists a nonconstant function $f : \mathbb{R} \to \mathbb{R}^+$ and an $\mathbb{R}$-algorithm (φ, T) as in Definition 2.2; because f is not constant, there exist two real numbers $\lambda_X, \lambda_Y \in \mathbb{R}$ with $f(\lambda_X) > f(\lambda_Y)$. Choose any distribution $\mu_X \in \mathcal{M}_1(\mathbb{R})$ with $m_1(\mu_X) = \lambda_X$, and consider a sequence $X = (X_n)_{n\geq 1}$ marginally distributed according to μ_X. For $\varepsilon > 0$ and an independent sequence of Bernoulli random variables $(B_n)_{n\geq 1}$ with success probability $\mathbb{P}(B_n = 1) = 1 - \mathbb{P}(B_n = 0) = 1 - \varepsilon$, independent from any other source of randomness, the sequence $Y = (Y_n)_{n\geq 1}$ defined by

$$Y_n = B_n X_n + \frac{\lambda_Y - \lambda_X(1-\varepsilon)}{\varepsilon}(1 - B_n) \tag{2.1}$$

is such that $\mathbb{E}(Y_n) = \lambda_Y$. For any integer n we have $Y_n = X_n$ with arbitrarily large probability $1 - \varepsilon$, where ε can be chosen arbitrarily small, while λ_Y and λ_X are distinct and fixed; this construction is pivotal in all the proofs of this article.

Let us first give an informal description of the proof. We will compare the outputs of the algorithm for the two input sequences $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ and a common auxiliary variable U. Suppose first that the algorithm terminates after n steps when fed with the sequence $(X_n)_{n\geq 1}$. By tuning the value of ε we can make the events $\{(Y_1, \dots, Y_n) \neq (X_1, \dots, X_n)\}$ arbitrarily rare. On the other hand the expected outputs are set to $f(\lambda_X)$ for $(X_n)_{n\geq 1}$ and $f(\lambda_Y)$ for $(Y_n)_{n\geq 1}$, with $f(\lambda_Y) < f(\lambda_X)$. Hence, when the rare events $\{(Y_1, \dots, Y_n) \neq (X_1, \dots, X_n)\}$ do occur, the algorithm using $(Y_1, \dots, Y_n)$ needs to output a value sufficiently smaller than the value produced by the algorithm using $(X_1, \dots, X_n)$, so that the expected output can shift from $f(\lambda_X)$ to $f(\lambda_Y)$. However, the algorithm is not allowed to produce negative values so that the minimum output is zero. This would lead to a contradiction when the events $\{(Y_1, \dots, Y_n) \neq (X_1, \dots, X_n)\}$ are rare enough.

More formally denote by μ_Y the marginal law of each Y_n, namely

$$\mu_Y(dy) = (1-\varepsilon)\mu_X(dy) + \varepsilon 1_{\varepsilon^{-1}(\lambda_Y - \lambda_X(1-\varepsilon))}(dy).$$

The joint law on $([0,1], \mathbb{R}^{\mathbb{N}}, \mathbb{R}^{\mathbb{N}})$ of the random variables $(U, (X_n)_{n\geq 1}, (Y_n)_{n\geq 1})$ is denoted by $\breve{\mu}$; the marginal of $\breve{\mu}$ on its first two arguments is $(\text{Uniform}(0,1), \mu_X^{\otimes\mathbb{N}})$, and the marginal on its first and third arguments is $(\text{Uniform}(0,1), \mu_Y^{\otimes\mathbb{N}})$. We denote by $\breve{\mathbb{E}}$ the expectation with respect to $\breve{\mu}$ and by $\mathbb{E}_{U,X}$ and $\mathbb{E}_{U,Y}$ the expectations with respect to those two marginals, respectively.

Recall that the stopping times

$$\tau_X = \inf\{n : T_n(U, X_1, \dots, X_n) = 1\}, \qquad \tau_Y = \inf\{n : T_n(U, Y_1, \dots, Y_n) = 1\}$$

are by assumption almost surely finite and

$$\mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X})\big) = f(\lambda_X), \qquad \mathbb{E}_{U,Y}\big(\varphi_{\tau_Y}(U, Y_1, \ldots, Y_{\tau_Y})\big) = f(\lambda_Y).$$

Notice further that

$$\tau_X 1_{L_n \cap M_n} = \tau_Y 1_{L_n \cap M_n},$$

where we have defined the sets $L_n = \{\omega : \tau_X \leq n\}$ and $M_n = \{\omega : B_1 = \cdots = B_n = 1\} \subseteq \{\omega : X_1 = Y_1, \ldots, X_n = Y_n\}$. Since φ_{τ_Y} is almost surely nonnegative, we have for all $n \geq 1$,

$$\begin{aligned} \mathbb{E}_{U,Y}\big(\varphi_{\tau_Y}(U, Y_1, \ldots, Y_{\tau_Y})\big) &= \check{\mathbb{E}}\big(\varphi_{\tau_Y}(U, Y_1, \ldots, Y_{\tau_Y})\big) \\ &\geq \check{\mathbb{E}}\big(\varphi_{\tau_Y}(U, Y_1, \ldots, Y_{\tau_Y}) 1_{L_n \cap M_n}\big) \\ &= \check{\mathbb{E}}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n \cap M_n}\big). \end{aligned} \tag{2.2}$$

The random variables $(B_n)_{n \geq 1}$ are independent of any other source of randomness so that for all $n \geq 1$, we have

$$\begin{aligned} &\check{\mathbb{E}}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n \cap M_n}\big) \\ &\quad = (1-\varepsilon)^n \check{\mathbb{E}}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n}\big) \\ &\quad = (1-\varepsilon)^n \mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n}\big). \end{aligned} \tag{2.3}$$

The dominated convergence theorem yields

$$\begin{aligned} \lim_{n \to \infty} \mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n}\big) &= \mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X})\big) \\ &= f(\lambda_X) \end{aligned}$$

so that for any $\delta > 0$, there exists $n_0 = n_0(\delta) \in \mathbb{N}$ such that for all $n \geq n_0$,

$$f(\lambda_X) - \delta \leq \mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_n}\big) \leq f(\lambda_X). \tag{2.4}$$

One can choose $\delta > 0$ and $\eta > 0$ such that $f(\lambda_Y) + \eta < f(\lambda_X) - \delta$. Equations (2.2), (2.3) and (2.4) yield that for some integer $n_0 = n_0(\delta)$ and any $\varepsilon > 0$, we have

$$\begin{aligned} f(\lambda_Y) &= \mathbb{E}_{U,Y}\big(\varphi_{\tau_Y}(U, Y_1, \ldots, Y_{\tau_Y})\big) \geq \check{\mathbb{E}}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_{n_0} \cap M_{n_0}}\big) \\ &= (1-\varepsilon)^{n_0} \mathbb{E}_{U,X}\big(\varphi_{\tau_X}(U, X_1, \ldots, X_{\tau_X}) 1_{L_{n_0}}\big) \\ &\geq (1-\varepsilon)^{n_0}\big(f(\lambda_X) - \delta\big) > (1-\varepsilon)^{n_0}\big(f(\lambda_Y) + \eta\big). \end{aligned}$$

We obtain a contradiction for $\varepsilon > 0$ small enough. □

Theorem 2.1 indicates in particular that given U-estimators $(X_n)_{n \geq 1}$ of a quantity λ and without additional knowledge on these estimators, we cannot obtain U^+-estimators of neither $\exp(\lambda)$ nor $1/\lambda$.

Another question of interest arises in the case where $\mathcal{X} = \mathbb{R}$, we are given U-estimators of a quantity $\lambda > 0$ and we want to construct a U^+-estimator Y of the same quantity λ. This is not exactly equivalent to asking whether there exists an f-factory for $f : x \mapsto x$, first because we have only defined f-factories for f taking values in $\mathbb{R}^+$, and second because in Definition 2.2 the algorithm should work for any variable distributed as $\pi \in \mathcal{M}_1(\mathbb{R})$, whereas here we only consider distributions with expectation in $\mathbb{R}^+$.

LEMMA 2.1. *Let $\eta \geq 0$ be a known constant. There does not exist an $\mathbb{R}$-algorithm $\mathcal{A} \equiv (\varphi, T)$ such that for any independent sequence $X = (X_n)_{n\geq 1}$ marginally distributed as $\pi \in \mathcal{M}_1(\mathbb{R})$ with $m_1(\pi) > \eta$ and an auxiliary random variable $U \sim \text{Uniform}(0, 1)$ independent from $(X_n)_{n\geq 1}$, the random variable $Y = \mathcal{A}(U, X)$ is a nonnegative unbiased estimator of $m_1(\pi)$.*

PROOF. We follow the same arguments as in the proof of Theorem 2.1. Consider $\lambda_X, \lambda_Y \in \mathbb{R}^+$ with $\lambda_X > \lambda_Y > \eta$, and an algorithm $\mathcal{A} \equiv (\varphi, T)$ as in the statement of Lemma 2.1. Let $\mu_X \in \mathcal{M}_1(\mathbb{R})$ with $m_1(\mu_X) = \lambda_X$, and consider an sequence $X = (X_n)_{n\geq 1}$ marginally distributed according to μ_X. One can define Y as in equation (2.1). Since $\mathbb{E}(Y) = \lambda_Y \geq 0$, one can construct the same contradiction as in the proof of Theorem 2.1. □

The presence of $\eta \geq 0$ in the statement might seem cumbersome but emphasizes that the contradiction does not stem from distributions with expectation arbitrarily close to zero. According to the Lemma 2.1, even if one knows that a sequence of estimators has expectation larger than one, say, it is still impossible to design an algorithm transforming that sequence into a nonnegative random variable with the same expectation.

In the light of the nonexistence of f-factories when $\mathcal{X} = \mathbb{R}$, as stated in Theorem 2.1, we propose to study their existence when $\mathcal{X}$ is a subset of $\mathbb{R}$ in the next section.

3. Existence under stronger assumptions.

3.1. *Case where $\mathcal{X} = [a, +\infty)$ or $\mathcal{X} = (-\infty, b]$.*

LEMMA 3.1. *Let $a, b \in \mathbb{R}$ be two real numbers:*

- *For an f-factory to exist with $\mathcal{X} = [a, \infty)$ and $f : \mathcal{X} \to \mathbb{R}^+$, f must be increasing.*
- *For a g-factory to exist with $\mathcal{X} = (-\infty, b]$ and $g : \mathcal{X} \to \mathbb{R}^+$, g must be decreasing.*

PROOF. By symmetry we prove only the first assertion. For the sake of contradiction assume that there exist $a \leq \lambda_X < \lambda_Y$ with $f(\lambda_X) > f(\lambda_Y)$ and an algorithm $\mathcal{A} \equiv (\varphi, T)$ as in Definition 2.2. Choose any distribution $\mu_X \in \mathcal{M}_1([a, \infty))$ with $m_1(\mu_X) = \lambda_X$ and an independent sequence $X = (X_n)_{n\geq 1}$ marginally distributed according to μ_X. For $\varepsilon \in (0, 1)$, consider the sequence $Y = (Y_n)_{n\geq 1}$ as defined in equation (2.1). For $\varepsilon > 0$ small enough we have $\mathcal{D}(Y) \in \mathcal{M}_1([a, \infty))$ since $\lambda_Y > \lambda_X$. One can then construct exactly the same contradiction as in the proof of Theorem 2.1. □

Lemma 3.1 indicates in particular that it is impossible to obtain U^+-estimators of $1/\lambda$ given U^+-estimators of a quantity $\lambda > 0$ without exploiting any other additional information on the distribution of these U^+-estimators. For $\mathcal{X} = [a, \infty)$ and some increasing functions f, there can be explicit constructions of f-factories. For example, there exists an f-factory for any function $f : [a, \infty) \to \mathbb{R}^+$ that can be expressed as a power series of the type

$$f(x) = \sum_{n=0}^{\infty} c_n (x - a)^n \qquad \text{with } c_n \geq 0 \text{ for all } n \geq 0. \tag{3.1}$$

Indeed, introduce an independent sequence of random variables $(X_n)_{n\geq 1}$ marginally distributed as $\mu_X \in \mathcal{M}_1([a, \infty))$ and an integer-valued random variable N; setting the weights $w_n = 1/\mathbb{P}(N \geq n)$ as in Section 1.2, Tonelli's theorem yields that the estimator

$$Y = \sum_{n=0}^{N} w_n c_n \prod_{k=1}^{n} (X_k - a),$$

where the product is equal to 1 when $n = 0$, is well defined, is almost surely nonnegative and has expectation $f(m_1(\mu_X))$.

The above discussion gives a construction of a *Poisson estimator*, that is, a U^+-estimator of $\lambda = \exp(\mathbb{E}[X])$ given a stream $(X_n)_{n\geq 1}$ of i.i.d. $[a, +\infty)$-valued random variables distributed as X. Indeed the exponential function can be expressed as in equation (3.1) with $c_n = \exp(a)/n!$. One can readily check that if X has a finite variance and if the random variable N does not decay too rapidly to zero, for instance, $\mathbb{P}(N \geq n) \geq C/(1 + \varepsilon)^n$ for some constants $C, \varepsilon > 0$ as is the case for a geometric random variable, then equation (1.3) holds with

$$S_n = \exp(a) + \sum_{k=1}^{n} \frac{\exp(a)}{k!} \prod_{j=1}^{k} (X_j - a)$$

and $S = S_\infty$. The resulting Poisson estimator is unbiased and has a finite variance.

For increasing functions in general, the existence of f-factories remains an open question. Denoting by $\mathcal{F}$ the class of functions of the form described by equation (3.1), and by $\mathcal{C}$ the class of functions $f : [a, +\infty) \to \mathbb{R}^+$ for which an

f-factory exists, the previous discussion shows that $\mathcal{F} \subset \mathcal{C}$, and we conjecture $\mathcal{F} = \mathcal{C}$. For f and g in $\mathcal{C}$, then $f + g$ and $f \times g$ are in $\mathcal{C}$. In the special case $a = 0$, then $f \circ g$ is also in $\mathcal{C}$. A random truncation argument also shows that if $h : [a, +\infty) \to \mathbb{R}^+$ can be expressed as the infinite sum $h = \sum_{k \geq 0} f_k$ for functions $f_k \in \mathcal{C}$, then $h \in \mathcal{C}$. The set of functions $\mathcal{F}$ is the smallest class of functions that contains positive constants and the function $x \mapsto (x - a)$ and that is stable by the above-described operations. Those operations leave $\mathcal{C}$ stable because of simple properties of the expectation, such as linearity and the identity $\mathbb{E}[X \times Y] = \mathbb{E}[X] \times \mathbb{E}[Y]$ for X independent from Y. Our conjecture is based on our inability to exploit other properties of the expectation to find functions that would be in $\mathcal{C}$ but not in $\mathcal{F}$.

3.2. *Case where $\mathcal{X} = [a, b]$.* The case of a bounded interval $\mathcal{X} = [a, b]$ is the most related to the Bernoulli factory described in Section 1.2. We highlight in this section the similarities and differences between the construction of nonnegative estimators and Bernoulli factories. We then give a complete characterization of functions $f : \mathcal{X} = [a, b] \to \mathbb{R}^+$ for which f-factories exist.

Arguments similar to the proof of Theorem 2.1 show that for an f-factory to exist, the function $f : \mathcal{X} \to \mathbb{R}^+$ has to be continuous. Such a function $f : \mathcal{X} \to \mathbb{R}^+$ is thus necessarily bounded, and we consider a nontrivial interval $[0, \gamma]$ containing its range. If a Bernoulli factory exists for the function $g : [0, 1] \to [0, 1]$ with $g(x) = f(a(1 - x) + bx)/\gamma$, then there exists an f-factory. Indeed, consider an i.i.d. sequence $X = (X_n)_{n \geq 1}$ marginally distributed according to $\mu_X \in \mathcal{M}_1(\mathcal{X})$. Introduce random variables $(B_n)_{n \geq 1}$, with $B_n := 1_{U_n \leq (X_n - a)/(b - a)}$ where $(U_n)_{n \geq 1}$ is an i.i.d. sequence of random variables uniformly distributed on $(0, 1)$. Then $(B_n)_{n \geq 1}$ forms an i.i.d. sequence of Bernoulli random variables with mean $(m_1(\mu_X) - a)/(b - a)$. Therefore the Bernoulli factory for g takes the sequence $(B_n)_{n \geq 1}$ as input and produces a Bernoulli random variable $\widetilde{B}$ with mean $g((m_1(\mu_X) - a)/(b - a)) = f(m_1(\mu_X))/\gamma$. The random variable $\gamma \widetilde{B}$ is thus a nonnegative unbiased estimator of $m_1(\mu_X)$. As proved in Keane and O'Brien (1994), a necessary and sufficient condition on $g : [0, 1] \to [0, 1]$ for the existence of a Bernoulli factory is

$$\exists \varepsilon > 0, \exists n \in \mathbb{N}, \forall x \in [0, 1] \qquad \min\bigl(g(x), 1 - g(x)\bigr) \geq \varepsilon \min\bigl(x^n, (1 - x)^n\bigr).$$

It follows that an f-factory exists as soon as the condition $\min(f(x), \gamma - f(x)) \geq \varepsilon \min((x - a)^n, (b - x)^n)$ is satisfied for some $\varepsilon > 0$, $n \in \mathbb{N}$ and all $x \in [a, b]$. Theorem 3.1 shows in fact that

$$\exists \varepsilon > 0, \exists n \in \mathbb{N}, \forall x \in [a, b] \qquad f(x) \geq \varepsilon \min\bigl((x - a)^n, (b - x)^n\bigr) \tag{3.2}$$

is a necessary and sufficient condition for an f-factory to exist. The necessary condition $1 - g(x) \geq \varepsilon \min(x^n, (1 - x)^n)$ for the Bernoulli factory problem to have a solution comes from the fact that the Bernoulli factory has to produce a $\{0, 1\}$-valued estimator; we only need to construct a $[0, \infty)$-valued estimator and can thus get away with the weaker condition (3.2).

THEOREM 3.1. *Let $\mathcal{X} = [a, b]$ be a real interval and $f : \mathcal{X} \to \mathbb{R}^+$ a continuous function that is not identically zero. There exists an f-factory if and only if condition (3.2) holds.*

PROOF. The sufficiency is proved as a consequence of the results proved in Keane and O'Brien (1994). The proof of the necessity requires different arguments.

Sufficiency. Let $f : \mathcal{X} \to \mathbb{R}^+$ be a continuous function that satisfies condition (3.2). Since f is bounded on $\mathcal{X}$, one can find $\gamma \geq \max_{x \in \mathcal{X}} f(x)$ large enough such that $\gamma - f(x) > \varepsilon \min((x-a)^n, (b-x)^n)$ for all $x \in \mathcal{X}$. The discussion before the statement of Theorem 3.1 thus shows that an f-factory can be constructed.

Necessity. For notational convenience, we present the proof in the case $\mathcal{X} = [0, 1]$. The general case $\mathcal{X} = [a, b]$ is identical. Let $\mathcal{A} \equiv (T, \varphi)$ be an f-factory for some function $f : [0, 1] \to \mathbb{R}^+$. For $x_{1:n} = (x_1, \ldots, x_n) \in \{0, 1\}^n$ and a random variable U uniformly distributed on $(0, 1)$, we denote by $F_n(x_{1:n})$ the set of events such that the algorithm terminates after having processed $x_{1:n}$, that is,

$$F_n(x_{1:n}) = \{\omega : \inf\{1 \leq k \leq n : T_k(U, x_1, \ldots, x_k) = 1\} = n\}$$

with the convention $\inf\{\varnothing\} = \infty$. We define the expected output given $x_{1:n}$ by

$$\Psi_n(x_{1:n}) = \mathbb{E}\big(1_{F_n(x_{1:n})} \varphi_n(U, x_1, \ldots, x_n)\big).$$

For any index $n \geq 1$ and $x_{1:n} \in \{0, 1\}^n$, $\Psi_n(x_{1:n})$ is a nonnegative real number. By Definition 2.2 for any $z \in [0, 1]$ and an i.i.d. sequence $(X_n)_{n \geq 1}$ of Bernoulli random variables with mean $z \in [0, 1]$, we have

$$f(z) = \mathbb{E}\left(\sum_{n=1}^{\infty} \Psi_n(X_{1:n})\right) = \sum_{n=1}^{\infty} \sum_{x_{1:n} \in \{0,1\}^n} \mathbb{P}(X_{1:n} = x_{1:n}) \Psi_n(x_{1:n}).$$

For any index $n \geq 1$ and $x_{1:n} \in \{0, 1\}^n$, defining $r = r(x_{1:n}) = \mathrm{Card}\{1 \leq i \leq n : x_i = 1\}$, we have $\mathbb{P}(X_{1:n} = x_{1:n}) = z^r (1 - z)^{n-r}$, and the above double sum can be written as

$$f(z) = \sum_{n=1}^{\infty} \sum_{x_{1:n} \in \{0,1\}^n} z^r (1 - z)^{n-r} \Psi_n(x_{1:n}) = \sum_{p,q \in \mathbb{N}^2} c_{p,q} z^p (1 - z)^q$$

for some nonnegative coefficient $c_{p,q} \geq 0$. Condition (3.2) follows. □

By Theorem 3.1 it is possible to obtain U^+-estimators of $\exp(\lambda)$ or $1/\lambda$ given U-estimators of λ with support in some known interval $[a, b]$. Indeed, for the exponential case, one can use either a Bernoulli factory or the Poisson estimator described at the end of Section 3.1. For the inverse case on a segment $[a, b] \subset (0, \infty)$,

one can use either a Bernoulli factory or a random truncation argument to the series expansion

$$\frac{1}{x} = \frac{1}{b}\sum_{k=0}^{\infty}\left(\frac{b-x}{b}\right)^k$$

to construct an unbiased estimate of $\lambda = 1/\mathbb{E}[X]$ given a stream $(X_n)_{n\geq 1}$ of i.i.d. $[a, b]$-valued random variables distributed as X.

4. Discussion.

4.1. *Summary of the analysis.* The results of Section 2.2 show that, for a nonconstant function $f : \mathbb{R} \to \mathbb{R}^+$, the ability to sample an unbiased estimator X of a quantity λ is not enough to obtain a nonnegative unbiased estimator of $f(\lambda)$. However, as described in Section 3, when additional information such as almost sure lower or upper bounds on X is available, an f-factory might exist. The case where f is increasing and the support of X is $[a, \infty)$ remains partly unsettled.

We have prescribed as input of f-factories unbiased estimators of arbitrary quantities $\lambda \in \mathbb{R}$; other types of input could be envisioned, such as estimators consistent in L_2. However, in this case we could first apply a debiasing technique recalled in Section 1.2 and then feed the output to an f-factory, and hence the conclusion would be similar. Finally we have not considered the multi-dimensional case $f : \mathbb{R}^d \to \mathbb{R}^+$ for $d > 1$ since, in the context of exact inference, quantities of interest are posterior density evaluations.

4.2. *Exact or inexact inference.* An advantage of exact methods, where no systematic bias remains, is that the error is entirely due to the variation in the Monte Carlo algorithm and thus is straightforward to quantify and to interpret [Wagner (1987)]. The trade-off between computational feasibility and exactness is ubiquitous in statistics, for instance, between Ensemble Kalman filters and particle filters [Frei and Künsch (2013)] or between approximate Bayesian computation and Markov chain Monte Carlo [Marin et al. (2012)]. In some contexts such as state space models, a nonnegative unbiased estimator of the likelihood can be directly obtained, and the pseudo-marginal approach is proven efficient [Andrieu, Doucet and Holenstein (2010)]. Our study indicates that in some contexts nonnegative unbiased estimators cannot be obtained, and thus the pseudo-marginal approach cannot be applied. Exact inference could still be performed using signed unbiased estimators, as in the computational physics literature [Girolami et al. (2013), Lin, Liu and Sloan (2000), Troyer and Wiese (2005)].

In Section 3 the existence of f-factories has been studied under additional assumptions on the support of the input sequence. These assumptions are consistent with recent Monte Carlo methods for large datasets that take advantage of almost sure bounds to bypass the evaluation of the full likelihood [Bardenet, Doucet and

Holmes (2014), Maclaurin and Adams (2014)], leading to exact methods or inexact methods with a controlled error. There exist inexact methods with no control of the bias, which do not require almost sure bounds, such as some approximations of Metropolis–Hastings algorithms [Ceperley and Dewing (1999), Nicholls, Fox and Watt (2012)] or of Langevin diffusions [Ahn, Korattikara and Welling (2012), Chen, Fox and Guestrin (2014), Welling and Teh (2011)].

When f-factories exist as in Section 3, we have discussed implementable schemes based on the Bernoulli factory or on random truncations of infinite series. The algorithms considered in Definition 2.2 terminate with probability one, but the expected computational time is not necessarily finite. Hence even if the method could be applied in principle, its computational cost might prevent any practical implementation. The recent literature on Bernoulli factories has focused on characterizing algorithms that generate the desired output using as few input variables as possible [Flegal and Herbei (2012), Łatuszyński et al. (2011), Nacu and Peres (2005), Thomas and Blanchet (2011)], whereas Rhee and Glynn (2012), Rhee and Glynn (2013) carefully study the expected computational cost of debiasing techniques. The minimum computational cost of f-factories could be studied as well.

Acknowledgments. The first author gratefully acknowledges EPSRC for funding this research through grant EP/K009362/1. We are grateful to Mourad Sabilellah for stimulating discussions. We thank the Associate Editor and two anonymous referees for their comments, which helped improving both the presentation and the content of this article.

REFERENCES

AHN, S., KORATTIKARA, A. and WELLING, M. (2012). Bayesian posterior sampling via stochastic gradient Fisher scoring. In *Proceedings of the* 29*th International Conference on Machine Learning* (*ICML*-12) 1591–1598.

ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo (with discussion). *J. Roy. Statist. Soc. Ser. B* **72** 357–385.

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

BARDENET, R., DOUCET, A. and HOLMES, C. (2014). Towards scaling up Markov chain Monte Carlo: An adaptive subsampling approach. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*-14) 405–413.

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

BERGER, J. O., BERNARDO, J. M. and SUN, D. (2009). The formal definition of reference priors. *Ann. Statist.* **37** 905–938. MR2502655

BESKOS, A., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. *Bernoulli* **12** 1077–1098. MR2274855

BESKOS, A. and ROBERTS, G. O. (2005). Exact simulation of diffusions. *Ann. Appl. Probab.* **15** 2422–2444. MR2187299

BESKOS, A., PAPASPILIOPOULOS, O., ROBERTS, G. O. and FEARNHEAD, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 333–382. MR2278331

BHANOT, G. and KENNEDY, A. D. (1985). Bosonic lattice gauge theory with noise. *Phys. Lett. B* **157** 70–76.

CEPERLEY, D. M. and DEWING, M. (1999). The penalty method for random walks with uncertain energies. *J. Chem. Phys.* **110** 9812.

CHEN, T., FOX, E. B. and GUESTRIN, C. (2014). Stochastic gradient Hamiltonian Monte Carlo. In *Proceedings of the* 31*st International Conference on Machine Learning* (*ICML*-14) 1683–1691.

DEL MORAL, P., DOUCET, A. and JASRA, A. (2007). Sequential Monte Carlo for Bayesian computation. In *Bayesian Statistics* **8**: *Proceedings of the Eighth Valencia International Meeting*, *June* 2–6, 2006 (J. M. Bernardo, M. J. Bayarri, J. O. Degroot, A. P. Dawid, D. Heckerman, A. M. Smith and M. West, eds.) 115–148. Oxford Univ. Press, Oxford. MR2433191

EVERITT, R. G. (2012). Bayesian parameter estimation for latent Markov random fields and social networks. *J. Comput. Graph. Statist.* **21** 940–960. MR3005805

FEARNHEAD, P., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2008). Particle filters for partially observed diffusions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 755–777. MR2523903

FEARNHEAD, P., PAPASPILIOPOULOS, O., ROBERTS, G. O. and STUART, A. (2010). Random-weight particle filtering of continuous time processes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 497–512. MR2758525

FLEGAL, J. M. and HERBEI, R. (2012). Exact sampling for intractable probability distributions via a Bernoulli factory. *Electron. J. Stat.* **6** 10–37. MR2879671

FREI, M. and KÜNSCH, H. R. (2013). Bridging the ensemble Kalman and particle filters. *Biometrika* **100** 781–800. MR3142332

GIROLAMI, M., LYNE, A. M., STRATHMANN, H., SIMPSON, D. and ATCHADE, Y. (2013). Playing Russian roulette with intractable likelihoods. Preprint. Available at arXiv:1306.4032.

JOURDAIN, B. and SBAI, M. (2007). Exact retrospective Monte Carlo computation of arithmetic average Asian options. *Monte Carlo Methods Appl.* **13** 135–171. MR2338086

KEANE, M. S. and O'BRIEN, G. L. (1994). A Bernoulli factory. *ACM Trans. Model. Comput. Simul.* **4** 213–219.

KENNEDY, A. D. and KUTI, J. (1985). Noise without noise: A new Monte Carlo method. *Phys. Rev. Lett.* **54** 2473–2476.

KLEINER, A., TALWALKAR, A., SARKAR, P. and JORDAN, M. I. (2014). A scalable bootstrap for massive data. *J. Roy. Statist. Soc. Ser. B* **76** 795–816.

KUTI, J. (1982). Stochastic method for the numerical study of lattice fermions. *Phys. Rev. Lett.* **49** 183–186.

ŁATUSZYŃSKI, K., KOSMIDIS, I., PAPASPILIOPOULOS, O. and ROBERTS, G. O. (2011). Simulating events of unknown probabilities via reverse time martingales. *Random Structures Algorithms* **38** 441–452. MR2829311

LIN, L., LIU, K. F. and SLOAN, J. (2000). A noisy Monte Carlo algorithm. *Phys. Rev. D* **61** 074505.

LIU, J. S. and CHEN, R. (1998). Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93** 1032–1044. MR1649198

MACLAURIN, D. and ADAMS, R. P. (2014). Firefly Monte Carlo: Exact MCMC with subsets of data. Preprint. Available at arXiv:1403.5693.

MARIN, J.-M., PUDLO, P., ROBERT, C. P. and RYDER, R. J. (2012). Approximate Bayesian computational methods. *Stat. Comput.* **22** 1167–1180. MR2992292

MCLEISH, D. (2011). A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.* **17** 301–315. MR2890424

MØLLER, J., PETTITT, A. N., REEVES, R. and BERTHELSEN, K. K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93** 451–458. MR2278096

NACU, Ş. and PERES, Y. (2005). Fast simulation of new coins from old. *Ann. Appl. Probab.* **15** 93–115. MR2115037

NICHOLLS, G. K., FOX, C. and WATT, A. M. (2012). Coupled MCMC with a randomized acceptance probability. Preprint. Available at arXiv:1205.6857.

OLSSON, J. and STRÖJBY, J. (2011). Particle-based likelihood inference in partially observed diffusion processes using generalised Poisson estimators. *Electron. J. Stat.* **5** 1090–1122. MR2836770

PAPASPILIOPOULOS, O. (2011). A methodological framework for Monte Carlo probabilistic inference for diffusion processes. In *Bayesian Time Series Models* (D. Barber, A. T. Cemgil and S. Chiappa, eds.) 82–99. Cambridge Univ. Press, Cambridge. MR2894230

RHEE, C.-H. and GLYNN, P. W. (2012). A new approach to unbiased estimation for SDE's. In *Proceedings of the Winter Simulation Conference*. 17:1–17:7. Winter Simulation Conference, Berlin.

RHEE, C.-H. and GLYNN, P. W. (2013). Unbiased estimation with square root convergence for SDE models. Technical report, Stanford Univ., Stanford, CA.

RYCHLIK, T. (1990). Unbiased nonparametric estimation of the derivative of the mean. *Statist. Probab. Lett.* **10** 329–333. MR1069911

RYCHLIK, T. (1995). A class of unbiased kernel estimates of a probability density function. *Appl. Math.* (*Warsaw*) **22** 485–497. MR1314912

SERMAIDIS, G., PAPASPILIOPOULOS, O., ROBERTS, G. O., BESKOS, A. and FEARNHEAD, P. (2015). Markov chain Monte Carlo for exact inference for diffusions. *Scand. J. Stat.* **40** 294–321.

THOMAS, A. C. and BLANCHET, J. H. (2011). A practical implementation of the Bernoulli factory. Preprint. Available at arXiv:1106.2508.

TRAN, M. N., SCHARTH, M., PITT, M. K. and KOHN, R. (2013). Importance sampling squared for Bayesian inference in latent variable models Preprint. Available at arXiv:1309.3339.

TROYER, M. and WIESE, U.-J. (2005). Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Phys. Rev. Lett.* **94** 170201.

WAGNER, W. (1987). Unbiased Monte Carlo evaluation of certain functional integrals. *J. Comput. Phys.* **71** 21–33. MR0895521

WALKER, S. G. (2011). Posterior sampling when the normalizing constant is unknown. *Comm. Statist. Simulation Comput.* **40** 784–792. MR2783887

WELLING, M. and TEH, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the* 28*th International Conference on Machine Learning* (*ICML*-11) 681–688.

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1, SOUTH PARKS ROAD
OX1 3TG OXFORD
UNITED KINGDOM
E-MAIL: pierre.jacob@stats.ox.ac.uk

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
NATIONAL UNIVERSITY OF SINGAPORE
BLOCK S16, LEVEL 7, 6 SCIENCE DRIVE 2
SINGAPORE 117546
E-MAIL: a.h.thiery@nus.edu.sg