

# DOES WASTE RECYCLING REALLY IMPROVE THE MULTI-PROPOSAL METROPOLIS–HASTINGS ALGORITHM? AN ANALYSIS BASED ON CONTROL VARIATES

JEAN-FRANÇOIS DELMAS\* \*\* AND

BENJAMIN JOURDAIN,\* \*\*\* *Université Paris-Est*

## Abstract

The waste-recycling Monte Carlo (WRMC) algorithm introduced by physicists is a modification of the (multi-proposal) Metropolis–Hastings algorithm, which makes use of all the proposals in the empirical mean, whereas the standard (multi-proposal) Metropolis–Hastings algorithm uses only the accepted proposals. In this paper we extend the WRMC algorithm to a general control variate technique and exhibit the optimal choice of the control variate in terms of the asymptotic variance. We also give an example which shows that, in contradiction to the intuition of physicists, the WRMC algorithm can have an asymptotic variance larger than that of the Metropolis–Hastings algorithm. However, in the particular case of the Metropolis–Hastings algorithm called the Boltzmann algorithm, we prove that the WRMC algorithm is asymptotically better than the Metropolis–Hastings algorithm. This last property is also true for the multi-proposal Metropolis–Hastings algorithm. In this last framework we consider a linear parametric generalization of WRMC, and we propose an estimator of the explicit optimal parameter using the proposals.

*Keywords:* Metropolis–Hastings algorithm; multi-proposal algorithm; Monte Carlo Markov chain; variance reduction; control variates; ergodic theorem; central limit theorem

2000 Mathematics Subject Classification: Primary 60F05; 60J10; 60J22; 65C40; 82B80

## 1. Introduction

The Metropolis–Hastings algorithm is used to compute the expectation  $\langle \pi, f \rangle$  of a function  $f$  under a probability measure,  $\pi$ , difficult to simulate. It relies on the construction, by an appropriate acceptance/rejection procedure, of a Markov chain  $(X_k, k \geq 0)$  with transition kernel  $P$  such that  $\pi$  is reversible with respect to  $P$  and the quantity of interest  $\langle \pi, f \rangle$  is estimated by the empirical mean

$$I_n(f) = \frac{1}{n} \sum_{k=1}^n f(X_k).$$

Received 27 September 2007; revision received 15 October 2009.

\* Postal address: CERMICS, École des Ponts, 6–8 Avenue Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France.

\*\* Email address: delmas@cermics.enpc.fr

\*\*\* Email address: jourdain@cermics.enpc.fr

Research supported by the ANR program ADAP<sup>2</sup>MC.

We shall recall the well-known properties of this estimation (consistency, asymptotic normality) in what follows. In particular, the quality or precision of the algorithm is measured through the asymptotic variance of the estimator of  $\langle \pi, f \rangle$ .

The waste-recycling Monte Carlo (WRMC) algorithm, introduced by physicists, is a modification of the Metropolis–Hastings algorithm, which makes use of all the proposals in the empirical mean, whereas the standard Metropolis–Hastings algorithm uses only the accepted proposals. To the authors' knowledge, the WRMC algorithm was first introduced in 1977 by Ceperley, Chester, and Kalos (see [4, Equation (35), p. 3085]). Without any proof, they claimed that 'The advantage of using this form is that some information about unlikely moves appears in the final answer, and the variance is lowered'. It is commonly assumed among physicists and supported by most of the simulations that the WRMC algorithm is more efficient than the Metropolis–Hastings algorithm, that is, the estimation given by the WRMC algorithm is consistent and has a smaller asymptotic variance. Another way to speed up the Metropolis–Hastings algorithm could be to use multiple proposals at each step instead of only one. According to Frenkel [7], the waste recycling can be particularly useful for these algorithms where many states are rejected.

Our aim is to clarify the presentation of the WRMC algorithms with one proposal and with multiple proposals, and to present a first rigorous study of these algorithms. We will give in Section 2 an introduction to our results in the finite state space case. Our main new results are stated in Theorem 3.1, which is a first step towards the comparison of the asymptotic variances. We shall detail their consequences in the didactic Section 2 for

- the WRMC algorithm through Propositions 2.1 (consistency of the estimation), 2.2 (asymptotic normality), and 2.3 (a first partial answer to the initial question: Does waste recycling really improve the Metropolis–Hastings Monte Carlo algorithm?),
- the multi-proposal WRMC algorithm through Propositions 2.4 (consistency of the estimation and asymptotic normality) and 2.5 (a second partial answer to the initial question: Does waste recycling really improve the Metropolis–Hastings Monte Carlo algorithm?).

The study of the WRMC estimator in the form  $I_n(f) + J_n(f)$ , for a given functional  $J$ , leads us to rewrite the WRMC algorithm as a particular case of a general control variate problem by considering the estimators  $I_n(f) + J_n(\psi)$ , where the function  $\psi$  is possibly different from  $f$ . In the multi-proposal framework, the consistency (or convergence) of this general algorithm and its asymptotic normality are stated in Theorem 3.1 in Section 3. We also give its asymptotic variance and prove that the optimal choice of  $\psi$  in terms of asymptotic variance is the solution,  $F$ , of the Poisson equation (2.6). This choice achieves variance reduction, but the function  $F$  is difficult to compute. It is possible to replace it by an approximation. In some sense,  $f$  is such an approximation and, for this particular choice, we recover the waste-recycling estimator introduced by physicists. In Section 5, which is dedicated to the single-proposal case, we give a simple counterexample (see Subsection 5.2) which shows that the WRMC algorithm does not in general improve the Metropolis–Hastings algorithm: the WRMC algorithm can have an asymptotic variance larger than that of the Metropolis–Hastings algorithm. Athènes [3] also observed variance augmentation in some numerical computations of free energy. However, in the particular case of the Metropolis–Hastings algorithm called the Boltzmann algorithm, we prove in Section 4 that the (multi-proposal) WRMC algorithm is asymptotically better than the (multi-proposal) Metropolis–Hastings algorithm. In this particular framework we determine the optimal value  $b_\star$  of  $b$  for the parametric control variate  $J_n(bf)$ . This optimal value can be estimated using the Markov chain  $(X_k, 0 \leq k \leq n)$ .

### 2. Didactic version of the results

For simplicity, we assume in the present section that  $E$  is a finite set. Let  $\langle \nu, h \rangle = \sum_{x \in E} \nu(x)h(x)$  denote the ‘integration’ of a real function defined on  $E$ ,  $h = (h(x), x \in E)$ , with respect to a measure on  $E$ ,  $\nu = (\nu(x), x \in E)$ .

Let  $\pi$  be a probability measure on  $E$  such that  $\pi(x) > 0$  for all  $x \in E$ , and let  $f$  be a real function defined on  $E$ . The Metropolis–Hastings algorithm gives an estimation of  $\langle \pi, f \rangle$  as the almost-sure limit of the empirical mean of  $f$ ,  $(1/n) \sum_{k=1}^n f(X_k)$ , as  $n$  goes to infinity, where  $X = (X_n, n \geq 0)$  is a Markov chain which is reversible with respect to the probability measure  $\pi$ .

#### 2.1. The Metropolis–Hastings algorithm

The Markov chain  $X = (X_n, n \in \mathbb{N})$  of the Metropolis–Hastings algorithm is built in the following way. Let  $Q$  be an irreducible transition matrix over  $E$  such that, for all  $x, y \in E$ , if  $Q(x, y) = 0$  then  $Q(y, x) = 0$ . The transition matrix  $Q$  is called the selection matrix.

For  $x, y \in E$  such that  $Q(x, y) > 0$ , let  $(\rho(x, y), \rho(y, x)) \in (0, 1]^2$  be such that

$$\rho(x, y)\pi(x)Q(x, y) = \rho(y, x)\pi(y)Q(y, x). \tag{2.1}$$

The function  $\rho$  is viewed as an acceptance probability. For example, we obtain such a function  $\rho$  by setting

$$\rho(x, y) = \gamma \left( \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)} \right) \quad \text{for all } x, y \in E \text{ such that } Q(x, y) > 0, \tag{2.2}$$

where  $\gamma$  is a function with values in  $(0, 1]$  such that  $\gamma(u) = u\gamma(1/u)$ . Usually, we take  $\gamma(u) = \min(1, u)$  for the Metropolis algorithm. The case  $\gamma(u) = u/(1 + u)$  is known as the Boltzmann algorithm or Barker algorithm.

Let  $X_0$  be a random variable taking values in  $E$  with probability distribution  $\nu_0$ . At step  $n$ ,  $X_0, \dots, X_n$  are given. The proposal at step  $n + 1$ ,  $\tilde{X}_{n+1}$ , is distributed according to  $Q(X_n, \cdot)$ . This proposal is accepted with probability  $\rho(X_n, \tilde{X}_{n+1})$  and then  $X_{n+1} = \tilde{X}_{n+1}$ . If it is rejected then we set  $X_{n+1} = X_n$ .

It is easy to check that  $X = (X_n, n \geq 0)$  is a Markov chain with transition matrix  $P$  defined by

$$P(x, y) = \begin{cases} Q(x, y)\rho(x, y) & \text{if } x \neq y, \\ 1 - \sum_{z \neq x} P(x, z) & \text{if } x = y, \end{cases} \quad \text{for all } x, y \in E. \tag{2.3}$$

Furthermore,  $X$  is reversible with respect to the probability measure  $\pi$ :  $\pi(x)P(x, y) = \pi(y)P(y, x)$  for all  $x, y \in E$ . This property is also called detailed balance. By summation over  $y \in E$  we deduce that  $\pi$  is an invariant probability for  $P$  (i.e.  $\pi P = \pi$ ). The irreducibility of  $Q$  implies that  $P$  is irreducible. Since the probability measure  $\pi$  is invariant for  $P$ , we deduce that  $X$  is positive recurrent with (unique) invariant probability measure  $\pi$ . In particular, for any real-valued function  $f$  defined on  $E$ , the ergodic theorem (see, e.g. [9]) implies the consistency of the estimation:

$$\lim_{n \rightarrow \infty} I_n(f) = \langle \pi, f \rangle \quad \text{almost surely (a.s.),}$$

where

$$I_n(f) = \frac{1}{n} \sum_{k=1}^n f(X_k). \tag{2.4}$$

The asymptotic normality of the estimator  $I_n(f)$  is given by the following central limit theorem (see [6] or [9]):

$$\sqrt{n}(I_n(f) - \langle \pi, f \rangle) \xrightarrow{D} \mathcal{N}(0, \sigma(f)^2) \quad \text{as } n \rightarrow \infty.$$

Here  $\mathcal{N}(0, \sigma^2)$  denotes the Gaussian distribution with mean 0 and variance  $\sigma^2$ , the convergence holds in the distribution sense, and

$$\sigma(f)^2 = \langle \pi, F^2 \rangle - \langle \pi, (PF)^2 \rangle, \tag{2.5}$$

where  $F$  denotes the unique solution up to an additive constant of the Poisson equation

$$F(x) - PF(x) = f(x) - \langle \pi, f \rangle, \quad x \in E, \tag{2.6}$$

and  $Ph(x) = \sum_{y \in E} P(x, y)h(y)$ . Improving the Metropolis–Hastings algorithm means exhibiting other estimators of  $\langle \pi, f \rangle$  that are still consistent (i.e. estimators which converge a.s. to  $\langle \pi, f \rangle$ ) but have an asymptotic variance smaller than  $\sigma(f)^2$ .

**2.2. WRMC algorithm**

The classical estimation of  $\langle \pi, f \rangle$  by the empirical mean  $I_n(f)$  makes no use of the proposals  $\tilde{X}_k$  which have been rejected. For a long time, physicists have claimed that the efficiency of the estimation can be improved by including these rejected states in the sampling procedure. They suggest to use the so-called WRMC algorithm, which consists in replacing  $f(X_k)$  in  $I_n(f)$  by a weighted average of  $f(X_{k-1})$  and  $f(\tilde{X}_k)$ . For the natural choice of weights corresponding to the conditional expectation of  $f(X_k)$  with respect to  $(X_{k-1}, \tilde{X}_k)$ , we obtain the following estimator of  $\langle \pi, f \rangle$ :

$$\begin{aligned} I_n^{\text{WRMC}}(f) &= \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{E}[f(X_{k+1}) \mid X_k, \tilde{X}_{k+1}] \\ &= \frac{1}{n} \sum_{k=0}^{n-1} (\rho(X_k, \tilde{X}_{k+1})f(\tilde{X}_{k+1}) + (1 - \rho(X_k, \tilde{X}_{k+1}))f(X_k)). \end{aligned}$$

We shall study in Section 6.2 another choice for the weights also considered by Frenkel [8]. Note that the WRMC algorithm requires the evaluation of  $f$  for all the proposals whereas the Metropolis–Hastings algorithm evaluates  $f$  for only the accepted proposals. Other algorithms using all the proposals, such as the Rao–Blackwell Metropolis–Hastings algorithm, have been studied; see, for example, [12, Section 6.4.2] and the references therein, as well as [5] for a different approach. In the Rao–Blackwell Metropolis–Hastings algorithm, the weight of  $f(\tilde{X}_{k+1})$  depends on all the proposals  $\tilde{X}_1, \dots, \tilde{X}_n$ . It is thus necessary to keep in memory the values of all proposals in order to compute the estimation of  $\langle \pi, f \rangle$ .

We can easily check that  $I_n^{\text{WRMC}}(f) - I_n(f) = J_n(f)$ , where, for any real function  $\psi$  defined on  $E$ ,

$$\begin{aligned} J_n(\psi) &= \frac{1}{n} \sum_{k=0}^{n-1} (\mathbb{E}[\psi(X_{k+1}) \mid X_k, \tilde{X}_{k+1}] - \psi(X_{k+1})) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} (\rho(X_k, \tilde{X}_{k+1})\psi(\tilde{X}_{k+1}) + (1 - \rho(X_k, \tilde{X}_{k+1}))\psi(X_k) - \psi(X_{k+1})). \end{aligned}$$

Note that  $J_n(\psi) = 0$  when  $\psi$  is constant. We can consider a more general estimator of  $\langle \pi, f \rangle$  given by

$$I_n(f, \psi) = I_n(f) + J_n(\psi).$$

Note that  $I_n^{\text{WRMC}}(f) = I_n(f, f)$  and  $I_n(f) = I_n(f, 0)$ . It is easy to check that the bias of the estimator  $I_n(f, \psi)$  does not depend on  $\psi$ :  $E[I_n(f, \psi)] = E[I_n(f)]$ . Theorem 3.1, below, implies the following result on the estimator  $I_n(f, \psi)$ .

**Proposition 2.1.** *For any real functions  $\psi$  and  $f$  defined on  $E$ , the estimator  $I_n(f, \psi)$  of  $\langle \pi, f \rangle$  is consistent: a.s.,  $\lim_{n \rightarrow \infty} I_n(f, \psi) = \langle \pi, f \rangle$ .*

From this result,  $J_n(\psi)$  can be seen as a control variate and it is natural to look for  $\psi$  which minimizes the variance or the asymptotic variance of  $I_n(f, \psi)$ . Another class of control variates has been studied in [2] in the particular case of the independent Metropolis–Hastings algorithm, where  $Q(x \cdot)$  does not depend on  $x$ .

The last part of Theorem 3.1, below, implies the following result, in which we use Lemma 5.1, below, to derive the asymptotic variance expression. We shall write  $E_\pi$  when  $X_0$  is distributed under its invariant measure  $\pi$  (in particular,  $\langle \pi, f \rangle = E_\pi[f(X_0)]$ ).

**Proposition 2.2.** *For any real functions  $\psi$  and  $f$  defined on  $E$ , the estimator  $I_n(f, \psi)$  of  $\langle \pi, f \rangle$  is asymptotically normal:*

$$\sqrt{n}(I_n(f, \psi) - \langle \pi, f \rangle) \xrightarrow{D} \mathcal{N}(0, \sigma(f, \psi)^2) \quad \text{as } n \rightarrow \infty,$$

with asymptotic variance  $\sigma(f, \psi)^2$  given by

$$\begin{aligned} \sigma(f, \psi) = & \sigma(f)^2 - E_\pi[(1 - \rho(X_0, X_1))(F(X_1) - F(X_0))^2] \\ & + E_\pi[(1 - \rho(X_0, X_1))(\psi(X_1) - F(X_1) - \psi(X_0) + F(X_0))^2], \end{aligned}$$

where  $F$  solves the Poisson equation (2.6). In particular, for fixed  $f$ , the asymptotic variance  $\sigma(f, \psi)^2$  is minimal for  $\psi = F$  and this choice achieves variance reduction:  $\sigma(f, F)^2 \leq \sigma(f)^2$ .

Although optimal in terms of the asymptotic variance, the estimator  $I_n(f, F)$  is not for use in practice, since computing a solution of the Poisson equation is more complicated than computing  $\langle \pi, f \rangle$ . Nevertheless, the proposition suggests that using  $I_n(f, \psi)$ , where  $\psi$  is an approximation of  $F$ , might lead to a smaller asymptotic variance than in the standard Metropolis–Hastings algorithm. Some hint at the computation of an approximation of  $F$  by a Monte Carlo approach is, for instance, given in [10, p. 418–419]. Because of the series expansion  $F = \sum_{k \geq 0} P^k(f - \langle \pi, f \rangle)$ ,  $f$  can be seen as an approximation of  $F$  of order 0. Hence, the asymptotic variance of  $I_n^{\text{WRMC}}(f) = I_n(f, f)$  might be smaller than that of  $I_n(f)$  in some situations. It is a common belief in the physicist community (see [4] or [8]) that the inequality is always true. Note that, as remarked by Frenkel [8] in a particular case, the variance of each term of the sum in  $I_n^{\text{WRMC}}(f)$  is equal or smaller than the variance of each term of the sum in  $I_n(f)$  by Jensen’s inequality. But we also have to compare the covariance terms, which is not so obvious. We investigate whether the asymptotic variance of the WRMC algorithm is smaller than that of the standard Metropolis algorithm and reach the following conclusion that contradicts the intuition.

**Proposition 2.3.** (i) *In the Metropolis case, that is, when (2.2) holds with  $\gamma(u) = \min(1, u)$ , then it may happen that  $\sigma(f, f)^2 > \sigma(f)^2$ .*

(ii) When (2.2) holds with  $\gamma(u) = \alpha u / (1 + u)$  for some  $\alpha \in (0, 2)$ , then we have  $\sigma(f, f)^2 \leq \sigma(f)^2$ . Furthermore, for  $f$  nonconstant, the function  $b \mapsto \sigma(f, bf)^2$  is minimal at

$$b_\star = \frac{\langle \pi, f^2 \rangle - \langle \pi, f \rangle^2}{\langle \pi, f^2 - fPf \rangle} \tag{2.7}$$

and  $b_\star \geq 1/\alpha$ . When  $\alpha = 1$ , and if, moreover,  $\sigma(f, f)^2 > 0$ , then  $b_\star > 1$ .

**Remark 2.1.** Assume that  $f$  is not constant. The optimal parameter  $b_\star$  defined in (2.7) can be estimated by

$$\hat{b}_n = \frac{I_n(f^2) - I_n(f)^2}{I_n(f^2) - (1/n) \sum_{k=1}^n f(X_{k-1})f(X_k)}.$$

Note that, a.s.,  $\lim_{n \rightarrow \infty} \hat{b}_n = b_\star$  thanks to the ergodic theorem. Using Slutsky’s theorem, we can deduce from Proposition 2.2 that  $I_n(f) + \hat{b}_n J_n(f) = I_n(f, \hat{b}_n f)$  is an asymptotically normal estimator of  $\langle \pi, f \rangle$  with asymptotic variance  $\sigma(f, b_\star f)^2$ . Thus, in the framework of Proposition 2.3(ii), using the control variate  $\hat{b}_n J_n(f)$  strictly improves the WRMC estimator as soon as either  $\alpha < 1$  or  $\alpha = 1$  (Boltzmann algorithm) and  $\sigma(f, f)^2$  is positive. Note that when its asymptotic variance  $\sigma(f, f)^2$  is 0, then the WRMC estimator  $I_n^{\text{WRMC}}(f) = I_n(f, f)$  is equal to  $\langle \pi, f \rangle$ .

To prove assertion (i), we give an explicit counterexample such that  $\sigma(f, f)^2 > \sigma(f)^2$  in the Metropolis case (see Section 5.2 and (5.3), below). Assertion (ii) is also proved in Section 5 (see Proposition 5.1). Let us make some comments on its hypothesis, which holds with  $\alpha = 1$  for the Boltzmann acceptance rule.

- By (2.1) and since  $\rho(x, y)$  is an acceptance probability, the constant  $\alpha$  has to be smaller than

$$1 + \min_{x \neq y, Q(x,y) > 0} \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}.$$

- If there exists a constant  $c > 0$  such that, for all distinct  $x, y \in E$  such that  $Q(x, y) > 0$ , the quantity  $\pi(x)Q(x, y)/\pi(y)Q(y, x)$  is equal to  $c$  or  $1/c$  and (2.2) holds with  $\gamma$  such that  $\gamma(1/c) = \gamma(c)/c$ , then the hypothesis holds with  $\alpha = \gamma(c) + \gamma(1/c)$ . For example, assume that the transition matrix  $Q$  is symmetric and that  $\pi$  is written as a Gibbs distribution: for all  $x \in E$ ,  $\pi(x) = e^{-H(x)} / \sum_{y \in E} e^{-H(y)}$  for some energy function  $H$ . If the energy increases or decreases by the same amount  $\varepsilon$  for all the authorized transitions, then  $\pi(x)Q(x, y)/\pi(y)Q(y, x)$  is equal to  $c$  or  $1/c$  with  $c = e^\varepsilon$ .

According to [11], since, for all  $u > 0$ ,  $u/(1 + u) < \min(1, u)$ , in the absence of waste recycling, the asymptotic variance  $\sigma(f)^2$  is smaller in the Metropolis case than in the Boltzmann case for given  $\pi, Q$ , and  $f$ . So waste recycling always achieves variance reduction only for the worst choice of  $\gamma$ . Note, however, that the Boltzmann algorithm is used in the multi-proposal framework where we generalize our results.

**Remark 2.2.** When the computation of  $Pg$  is feasible for any function  $g: E \rightarrow \mathbb{R}$  (typically when, for every  $x \in E$ , the cardinality of  $\{y \in E: Q(x, y) > 0\}$  is small), then it is possible to use  $I_n(\psi - P\psi)$  as a control variate and approximate  $\langle \pi, f \rangle$  by  $I_n(f - (\psi - P\psi))$ . Since  $\pi$  is invariant with respect to  $P$ ,  $\langle \pi, \psi - P\psi \rangle = 0$  and, a.s.,  $I_n(f - (\psi - P\psi))$  converges to  $\langle \pi, f \rangle$  as  $n$  tends to  $\infty$ . Moreover, the asymptotic variance of the estimator is  $\sigma(f - \psi + P\psi)^2$ . Now,

remarking that

$$I_n(\psi - P\psi) = \frac{1}{n} \sum_{k=1}^n (\psi(X_k) - P\psi(X_{k-1})) + \frac{1}{n} (P\psi(X_0) - P\psi(X_n)) \tag{2.8}$$

we find that the bias difference

$$E[I_n(f - \psi + P\psi)] - E[I_n(f)] = \frac{1}{n} E[P\psi(X_0) - P\psi(X_n)]$$

is smaller than  $2 \max_{x \in E} |\psi(x)|/n$ .

For the choice  $\psi = F$ , this control variate is perfect, since according to (2.6), for each  $n \in \mathbb{N}^*$ ,  $I_n(f - (F - PF))$  is constant and equal to  $\langle \pi, f \rangle$ .

Setting  $f_0 = f - \langle \pi, f \rangle$ , we find that the asymptotic variance for the choice  $\psi = bf$  with  $b \in \mathbb{R}$  is

$$\begin{aligned} \sigma(f + b(Pf - f))^2 &= \langle \pi, (F - bf_0)^2 \rangle - \langle \pi, (PF - bPf_0)^2 \rangle \\ &= \langle \pi, F^2 - (PF)^2 \rangle - 2b\langle \pi, f_0(F - PF + PF - P^2F) \rangle \\ &\quad + b^2\langle \pi, f_0^2 - (Pf_0)^2 \rangle \\ &= \sigma(f)^2 - 2b\langle \pi, f_0(f_0 + Pf_0) \rangle + b^2\langle \pi, f_0^2 - (Pf_0)^2 \rangle, \end{aligned}$$

where we have used the Poisson equation (2.6) for the first and third equalities, and the reversibility of  $\pi$  with respect to  $P$  for the second equality. The parameter

$$\tilde{b} = \frac{\langle \pi, f_0(f_0 + Pf_0) \rangle}{\langle \pi, f_0^2 - (Pf_0)^2 \rangle} = \frac{\langle \pi, f(f + Pf) \rangle - 2\langle \pi, f \rangle^2}{\langle \pi, f^2 - (Pf)^2 \rangle},$$

which minimizes this variance, can be consistently estimated by

$$\frac{I_n(f(f + Pf)) - 2I_n(f)^2}{I_n(f^2 - (Pf)^2)}.$$

Note that the control variate  $J_n(\psi)$  is similar to  $I_n(\psi - P\psi)$  except that the conditional expectation  $P\psi(X_{k-1})$  of  $\psi(X_k)$  given  $X_{k-1}$  in the first term on the right-hand side of (2.8) is replaced by the conditional expectation of  $\psi(X_k)$  given  $(X_{k-1}, \tilde{X}_k)$ , which can always be easily computed. From this perspective, the minimality of the asymptotic variance of  $I_n(f, \psi)$  for  $\psi = F$  is not a surprise.

The comparison between  $\sigma(f, \psi)^2$  and  $\sigma(f - \psi + P\psi)^2$  can be deduced from Section 6.1, which is stated in the more general multi-proposal framework introduced in the next subsection. Note that the sign of  $\sigma(f, \psi)^2 - \sigma(f - \psi + P\psi)^2$  depends on  $\psi$ .

### 2.3. Multi-proposal WRMC algorithm

In the classical Metropolis–Hastings algorithm, there is only one proposal  $\tilde{X}_{n+1}$  at step  $n + 1$ . Around 1990, some extensions where only one state among multiple proposals is accepted were proposed in order to speed up the exploration of  $E$  (see [1] for a unifying presentation of MCMC algorithms including the multi-proposal Metropolis–Hastings algorithm). According to Frenkel [7], waste recycling can be particularly useful for these algorithms where many states are rejected.

To formalize these algorithms, we introduce a proposition kernel  $\mathcal{Q}: E \times \mathcal{P}(E) \rightarrow [0, 1]$ , where  $\mathcal{P}(E)$  denotes the set of parts of  $E$ , which describes how to randomly choose the set of proposals: for all  $x \in E$ ,

$$\mathcal{Q}(x, A) = 0 \quad \text{if } x \notin A \quad \text{and} \quad \sum_{A \in \mathcal{P}(E)} \mathcal{Q}(x, A) = 1. \tag{2.9}$$

The second condition says that  $\mathcal{Q}(x, \cdot)$  is a probability on  $\mathcal{P}(E)$ . The first condition ensures that the starting point is among the proposals. This last convention will allow us to transform the rejection/acceptance procedure into a selection procedure among the proposals.

The selection procedure is described by a probability  $\kappa$ . For  $(x, A) \in E \times \mathcal{P}(E)$ , let  $\kappa(x, A, \tilde{x}) \in [0, 1]$  denote the probability of choosing  $\tilde{x} \in A$  as the next state when the proposal set  $A$  has been chosen. We assume that  $\sum_{\tilde{x} \in A} \kappa(x, A, \tilde{x}) = 1$  (that is,  $\kappa(x, A, \cdot)$  is a probability measure) and that the following condition holds:

$$\pi(x)\mathcal{Q}(x, A)\kappa(x, A, \tilde{x}) = \pi(\tilde{x})\mathcal{Q}(\tilde{x}, A)\kappa(\tilde{x}, A, x) \quad \text{for all } A \in \mathcal{P}(E) \text{ and all } x, \tilde{x} \in A. \tag{2.10}$$

This condition is the analogue of (2.1) for a multi-proposal setting. For examples of nontrivial selection probabilities  $\kappa$ , see the text following Proposition 2.4, below.

The Markov chain  $X = (X_n, n \geq 0)$  is now defined inductively in the following way. Let  $X_0$  be a random variable taking values in  $E$  with probability distribution  $\nu_0$ . At step  $n$ ,  $X_0, \dots, X_n$  are given. The proposal set at step  $n + 1$ ,  $A_{n+1}$ , is distributed according to  $\mathcal{Q}(X_n, \cdot)$ . Then  $X_{n+1}$  is chosen distributed according to  $\kappa(X_n, A_{n+1}, \cdot)$ . It is easy to check that  $X$  is a Markov chain with transition matrix

$$P(x, y) = \sum_{\{A \in \mathcal{P}(E) : x, y \in A\}} \mathcal{Q}(x, A)\kappa(x, A, y).$$

Condition (2.10) ensures that  $X$  is reversible with respect to the probability measure

$$\pi : \pi(x)P(x, y) = \pi(y)P(y, x).$$

**Remark 2.3.** The multi-proposal Metropolis–Hastings algorithm generalizes the Metropolis–Hastings algorithm which can be recovered for the particular choice  $\mathcal{Q}(x, \{x, y\}) = Q(x, y)$  and, for  $y \neq x$ ,  $\kappa(x, \{x, y\}, y) = 1 - \kappa(x, \{x, y\}, x) = \rho(x, y)$ .

We keep the definition of  $I_n(f)$ , (2.4), but adapt the definitions of  $J_n(\psi)$  and  $I_n(f, \psi)$  as follows:

$$\begin{aligned} \mathcal{J}_n(\psi) &= \frac{1}{n} \sum_{k=0}^{n-1} (\mathbb{E}[\psi(X_{k+1}) \mid X_k, A_{k+1}] - \psi(X_{k+1})) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \left( \sum_{\tilde{x} \in A_{k+1}} \kappa(X_k, A_{k+1}, \tilde{x})\psi(\tilde{x}) - \psi(X_{k+1}) \right) \end{aligned} \tag{2.11}$$

and  $\mathcal{I}_n(f, \psi) = I_n(f) + \mathcal{J}_n(\psi)$ . The waste-recycling estimator of  $\langle \pi, f \rangle$  studied by Frenkel [7] is given by  $\mathcal{I}_n^{\text{WRMC}}(f) = \mathcal{I}_n(f, f)$ . Note that the bias of the estimator  $\mathcal{I}_n(f, \psi)$  does not depend on  $\psi$  (i.e.  $\mathbb{E}[\mathcal{I}_n(f, \psi)] = \mathbb{E}[I_n(f)]$ ). It turns out that Propositions 2.1 and 2.2 remain true in this multi-proposal framework (see Theorem 3.1, below) as soon as  $P$  is irreducible. Note that the irreducibility of  $P$  holds if and only if, for all  $x' \neq y \in E$ , there exist  $m \geq 1$



distinct  $x_0 = y, x_1, x_2, \dots, x_m = x' \in E$  and  $A_1, A_k, \dots, A_m \in \mathcal{P}(E)$  such that, for all  $k \in \{1, \dots, m\}, x_{k-1}, x_k \in A_k$  and

$$\prod_{k=1}^m \mathcal{Q}(x_{k-1}, A_k) \kappa(x_{k-1}, A_k, x_k) > 0. \tag{2.12}$$

**Proposition 2.4.** *Assume that  $P$  is irreducible. For any real functions  $\psi$  and  $f$  defined on  $E$ , the following assertions hold.*

- The estimator  $\mathfrak{I}_n(f, \psi)$  of  $\langle \pi, f \rangle$  is consistent: a.s.,  $\lim_{n \rightarrow \infty} \mathfrak{I}_n(f, \psi) = \langle \pi, f \rangle$ .
- The estimator  $\mathfrak{I}_n(f, \psi)$  of  $\langle \pi, f \rangle$  is asymptotically normal:

$$\sqrt{n}(\mathfrak{I}_n(f, \psi) - \langle \pi, f \rangle) \xrightarrow{D} \mathcal{N}(0, \sigma(f, \psi)^2) \text{ as } n \rightarrow \infty,$$

where the asymptotic variance (still denoted by)  $\sigma(f, \psi)^2$  is given by

$$\sigma(f, \psi)^2 = \sigma(f)^2 + \sum_{\substack{x \in E \\ A \in \mathcal{P}(E)}} \pi(x) \mathcal{Q}(x, A) [\text{var}_{\kappa_{x,A}}(\psi - F) - \text{var}_{\kappa_{x,A}}(F)],$$

with

$$\text{var}_{\kappa_{x,A}}(g) = \sum_{y \in A} \kappa(x, A, y) g(y)^2 - \left( \sum_{y \in A} \kappa(x, A, y) g(y) \right)^2.$$

- Moreover, for fixed  $f$ , the asymptotic variance  $\sigma(f, \psi)^2$  is minimal for  $\psi = F$ , where  $F$  solves the Poisson equation (2.6). In particular, this choice achieves variance reduction:  $\sigma(f, F)^2 \leq \sigma(f)^2$ .

We now give two examples of a nontrivial selection probability  $\kappa$  which satisfies condition (2.10). The first nontrivial selection probability,  $\kappa^M$ , defined by

$$\kappa^M(x, A, \tilde{x}) = \begin{cases} \frac{\pi(\tilde{x}) \mathcal{Q}(\tilde{x}, A)}{\max(\pi(\tilde{x}) \mathcal{Q}(\tilde{x}, A), \pi(x) \mathcal{Q}(x, A)) + \sum_{z \in A \setminus \{\tilde{x}\}} \pi(z) \mathcal{Q}(z, A)} & \text{if } \tilde{x} \neq x, \\ 1 - \sum_{z \in A \setminus \{x\}} \kappa^M(x, A, z) & \text{if } \tilde{x} = x, \end{cases} \tag{2.13}$$

generalizes the Metropolis selection given by (2.2) with  $\gamma(u) = \min(1, u)$ . (Note that, for  $x \neq \tilde{x}$ , we have  $\kappa^M(x, A, \tilde{x}) \leq \pi(\tilde{x}) \mathcal{Q}(\tilde{x}, A) / \sum_{z \in A \setminus \{x\}} \pi(z) \mathcal{Q}(z, A)$ , which implies that  $1 - \sum_{z \in A \setminus \{x\}} \kappa^M(x, A, z)$  is indeed nonnegative.) The second nontrivial selection probability,  $\kappa^B$ , which does not depend on the initial point  $x$ , and is defined by

$$\kappa^B(x, A, \tilde{x}) = \kappa^B(A, \tilde{x}) = \frac{\pi(\tilde{x}) \mathcal{Q}(\tilde{x}, A)}{\sum_{z \in A} \pi(z) \mathcal{Q}(z, A)}, \tag{2.14}$$

generalizes the Boltzmann (or Barker) selection given by (2.2) with  $\gamma(u) = u/(1 + u)$ . Note that, for both choices, the irreducibility condition (2.12) can be expressed only in terms of  $\mathcal{Q}$ :

$$\prod_{k=1}^m \mathcal{Q}(x_{k-1}, A_k) \mathcal{Q}(x_k, A_k) > 0.$$

For the selection probability in (2.14), we prove in Section 4 (see Proposition 4.1) that the waste recycling improves the Metropolis–Hastings algorithm.

**Proposition 2.5.** *When  $\kappa = \kappa^B$  is given by (2.14) (Boltzmann or Barker case), then we have  $\sigma(f, f)^2 \leq \sigma(f)^2$ . Furthermore, for  $f$  nonconstant, the function  $b \mapsto \sigma(f, bf)^2$  is minimal at  $b_\star$  defined by (2.7) and  $b_\star > 1$  when  $\sigma(f, f)^2 > 0$ .*

Note that the optimal value  $b_\star$  can be estimated by  $\hat{b}_n$ , which is computed using the proposals; see Remark 2.1. The control variate  $\hat{b}_n J_n(f)$  therefore improves the WRMC algorithm.

**2.4. Conclusion**

According to Proposition 2.3, the WRMC algorithm does not in general improve the Metropolis–Hastings algorithm. Nevertheless, even in the multi-proposal framework, it achieves variance reduction for the Boltzmann (or Barker) selection probability (see Proposition 2.5). In this case we improve it by computing the optimal multiplicative constant  $b^\star$  of the associated control variate and by providing a convergent estimator of  $b^\star$  (see Remark 2.1). Since the Metropolis selection probability is larger than the Boltzmann selection probability outside the diagonal (for  $\tilde{x} \neq x \in A$ ,  $\kappa^M(x, A, \tilde{x}) \geq \kappa^B(A, \tilde{x})$ ), according to [11], the asymptotic variance without any control variate is smaller in the Metropolis case than in the Boltzmann case. From formulae (2.13) and (2.14), it is likely that the difference decreases when the cardinality of the proposal sets increases. This is why we may expect the waste recycling to reduce the asymptotic variance even for the Metropolis selection probability in the multi-proposal framework when the cardinality of the proposal sets is large.

**3. Main result for general multi-proposal waste recycling**

Let  $(E, \mathcal{F}_E)$  be a measurable space such that  $\{x\} \in \mathcal{F}_E$  for all  $x \in E$ , and let  $\pi$  be a probability measure on  $E$ . Note that  $E$  is not assumed to be finite. Let  $\mathcal{P} = \{A \subset E; \text{card}(A) < \infty\}$  be the set of finite subsets of  $E$ . Let  $\bar{E} = \bigcup_{n \geq 1} E^n$ , and let  $\mathcal{F}_{\bar{E}}$  be the smallest  $\sigma$ -field on  $\bar{E}$  that contains  $A_1 \times \dots \times A_n$  for all  $A_i \in \mathcal{F}_E$  and  $n \geq 1$ . We consider the function  $\Gamma$  defined on  $\bar{E}$  and taking values on  $\mathcal{P}$  such that  $\Gamma((x_1, \dots, x_n))$  is the set  $\{x_1, \dots, x_n\}$  of distinct elements in  $(x_1, \dots, x_n)$ . We define  $\mathcal{F}_{\mathcal{P}}$ , a  $\sigma$ -field on  $\mathcal{P}$ , as the image of  $\mathcal{F}_{\bar{E}}$  by the application  $\Gamma$ . We consider a measurable proposition probability kernel  $\mathcal{Q}: E \times \mathcal{F}_{\mathcal{P}} \rightarrow [0, 1]$  such that

$$\int_{\mathcal{P}} \mathcal{Q}(x, dA) = 1 \quad \text{and} \quad \int_{\mathcal{P}} \mathcal{Q}(x, dA) \mathbf{1}_{\{x \notin A\}} = 0$$

(this is the analogue of (2.9)), and a measurable selection probability kernel  $\kappa: E \times \mathcal{P} \times \mathcal{F}_E \rightarrow [0, 1]$  such that, for  $x \in A$ , we have  $\kappa(x, A, A) = 1$ . Let  $\delta_y$  be the Dirac mass at point  $y$ . In particular, since  $A$  is finite, with a slight abuse of notation, we shall also write  $\kappa(x, A, dy) = \sum_{z \in A} \kappa(x, A, z) \delta_z(dy)$  and so  $\sum_{y \in A} \kappa(x, A, y) = 1$ .

We assume that the analogue of (2.10) holds, that is,

$$\pi(dx) \mathcal{Q}(x, dA) \kappa(x, A, dy) = \pi(dy) \mathcal{Q}(y, dA) \kappa(y, A, dx). \tag{3.1}$$

**Example 3.1.** We give the analogue of the Metropolis and Boltzmann selection kernels defined in (2.13) and (2.14) when  $E$  is finite. We consider  $N(dx, dA) = \pi(dx) \mathcal{Q}(x, dA)$  and a measure  $N_0(dA)$  on  $\mathcal{F}_{\mathcal{P}}$  such that  $\int_{x \in E} N(dx, dA)$  is absolutely continuous with respect to  $N_0(dA)$ . Since  $x \in A$  and  $A$  is finite  $N(dx, dA)$ -a.s., the decomposition of  $N$  with respect to  $N_0$  gives  $N(dx, dA) = N_0(dA) r_A(dx)$ , where  $r_A(dx) = \sum_{y \in A} r_A(y) \delta_y(dx)$  if  $A$  is finite and  $r_A(dx) = 0$  otherwise, and  $(x, A) \mapsto r_A(x)$  is jointly measurable.

The Metropolis selection kernel is given by, for  $x, y \in A$  and  $r_A \neq 0$ ,

$$\kappa^M(x, A, y) = \frac{r_A(y)}{\sum_{z \in A \setminus \{x, y\}} r_A(z) + \max(r_A(x), r_A(y))}, \tag{3.2}$$

if  $x \neq y$  and  $\kappa^M(x, A, x) = 1 - \sum_{y \in A \setminus \{x\}} \kappa^M(x, A, y)$ .

The Boltzmann selection kernel is given by, for  $x, y \in A$  and  $r_A \neq 0$ ,

$$\kappa^B(x, A, y) = \kappa^B(A, y) = \frac{r_A(y)}{\sum_{z \in A} r_A(z)}. \tag{3.3}$$

We choose these two selection kernels to be equal to the uniform distribution on  $A$  when  $r_A = 0$ . For these two selection kernels, (3.1) is satisfied.

**Example 3.2.** Let us give a natural example. Let  $\nu$  be a reference measure on  $E$  with no atoms, and let  $\pi$  be a probability measure on  $E$  with density with respect to  $\nu$ , which we still denote by  $\pi$ . Let  $\mathcal{Q}(x, \mathcal{A}) = P_x(\{x, Y_1, \dots, Y_n\} \subset \mathcal{A})$  for  $\mathcal{A} \in \mathcal{F}_{\mathcal{P}}$  be a selection procedure, where  $Y_1, \dots, Y_n$  are  $E$ -valued independent random variables with density with respect to  $\nu$  given by  $q(x, \cdot)$  under  $P_x$  and  $n \geq 1$  is fixed. We use the notation of Example 3.1. In this setting we choose  $N_0(dA) = \prod_{x \in A} \nu(dx)$  and the function  $r_A$  is given by, for  $x \in A$ ,  $r_A(x) = \pi(x) \prod_{z \in A \setminus \{x\}} q(x, z)$ .

The Markov chain  $X = (X_n, n \geq 0)$  is defined inductively in the following way. Let  $X_0$  be a random variable taking values in  $E$  with probability distribution  $\nu_0$ . At step  $n$ ,  $X_0, \dots, X_n$  are given. The proposal set at step  $n + 1$ ,  $A_{n+1}$ , is distributed according to  $\mathcal{Q}(X_n, \cdot)$ . Then  $X_{n+1}$  is chosen distributed according to  $\kappa(X_n, A_{n+1}, \cdot)$ . This is a particular case of the hit-and-run algorithm [1], where the proposal sets are always finite. It is easy to check that  $X$  is a Markov chain with transition kernel

$$P(x, dy) = \int_{\mathcal{P}} \mathcal{Q}(x, dA) \kappa(x, A, dy). \tag{3.4}$$

For  $f$  a real-valued measurable function defined on  $E$ , we shall write  $Pf(x)$  for  $\int_E P(x, dy) f(y)$  when this integral is well defined.

Condition (3.1) ensures that  $X$  is reversible with respect to

$$\pi : \pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

We also assume that  $X$  is Harris recurrent (see [9, Section 9]). This is equivalent to assuming that, for all  $B \in \mathcal{F}_E$  such that  $\pi(B) > 0$ , we have  $P(\text{card}\{n \geq 0; X_n \in B\} = \infty \mid X_0 = x) = 1$  for all  $x \in E$ .

**Example 3.3.** It is easy to check in Example 3.2 that  $X$  is Harris recurrent if the random walk with transition kernel  $q$  is itself Harris recurrent and, for all  $x \in E$ ,  $\mathcal{Q}(x, dA)$  a.s., for all  $y \in A$ ,  $\kappa(x, A, y) > 0$  (i.e.  $\int_{\mathcal{P}} \mathbf{1}_{\{\text{for all } y \in A, \kappa(x, A, y) > 0\}} \mathcal{Q}(x, dA) = 1$ ).

For  $f$  a real-valued measurable function defined on  $E$  and  $\nu$  a measure on  $E$ , we shall write  $\langle \nu, f \rangle$  for  $\int \nu(dy) f(y)$  when this integral is well defined.

Let  $f$  be a real-valued measurable function defined on  $E$  such that  $\langle \pi, |f| \rangle < \infty$ . Theorem 17.3.2 of [9] asserts that, a.s.,  $\lim_{n \rightarrow \infty} I_n(f) = \langle \pi, f \rangle$ , with  $I_n(f)$  defined by (2.4).

We consider the functional  $\mathcal{J}_n$  defined by

$$\begin{aligned} \mathcal{J}_n(\beta) &= \frac{1}{n} \sum_{k=0}^{n-1} (\mathbb{E}[\beta(X_k, A_{k+1}, X_{k+1}) \mid X_k, A_{k+1}] - \beta(X_k, A_{k+1}, X_{k+1})) \\ &= \frac{1}{n} \sum_{k=0}^{n-1} \left( \sum_{\tilde{x} \in A_{k+1}} \kappa(X_k, A_{k+1}, \tilde{x}) \beta(X_k, A_{k+1}, \tilde{x}) - \beta(X_k, A_{k+1}, X_{k+1}) \right) \end{aligned} \tag{3.5}$$

for  $\beta$  any real-valued measurable function defined on  $E \times \mathcal{P} \times E$ . We set  $\mathcal{J}_n(f, \beta) = I_n(f) + \mathcal{J}_n(\beta)$ . To prove the convergence and the asymptotic normality of the estimator  $\mathcal{J}_n(f, \beta)$  of  $\langle \pi, f \rangle$ , we shall use a martingale approach. In particular, we shall assume that there exists a solution  $F$  to the Poisson equation  $F - PF = f - \langle \pi, f \rangle$  such that  $\langle \pi, F^2 \rangle < \infty$  (see Theorem 17.4.2 and Condition (V.3) of [9] to ensure the existence of such a solution).

We introduce the following convenient notation. For a probability measure  $\nu$  on  $E$  and real-valued functions  $h$  and  $g$  defined on  $E$ , we respectively write, when well defined,

$$\text{cov}_\nu(h, g) = \langle \nu, gh \rangle - \langle \nu, g \rangle \langle \nu, h \rangle \quad \text{and} \quad \text{var}_\nu(h) = \langle \nu, h^2 \rangle - \langle \nu, h \rangle^2$$

for the covariance of  $g$  and  $h$ , and the variance of  $h$  with respect to  $\nu$ . We also write  $\kappa_{x,A}(\text{d}y)$  for the probability measure  $\kappa(x, A, \text{d}y)$  and  $\beta_{x,A}(\cdot)$  for the function  $\beta(x, A, \cdot)$ .

**Theorem 3.1.** *We assume that  $X$  is Harris recurrent, that  $\langle \pi, f^2 \rangle < \infty$ , that there exists a solution  $F$  to the Poisson equation  $F - PF = f - \langle \pi, f \rangle$  such that  $\langle \pi, F^2 \rangle < \infty$ , and that  $\beta$  is square integrable:  $\int \pi(\text{d}x) \mathcal{Q}(x, \text{d}A) \kappa(x, A, \text{d}y) \beta(x, A, y)^2 < \infty$ . Under these assumptions, the following assertions hold.*

- (i) *The estimator  $\mathcal{J}_n(f, \beta)$  of  $\langle \pi, f \rangle$  is consistent: a.s.,  $\lim_{n \rightarrow \infty} \mathcal{J}_n(f, \beta) = \langle \pi, f \rangle$ .*
- (ii) *The estimator  $\mathcal{J}_n(f, \beta)$  of  $\langle \pi, f \rangle$  is asymptotically normal:*

$$\sqrt{n}(\mathcal{J}_n(f, \beta) - \langle \pi, f \rangle) \xrightarrow{\text{D}} \mathcal{N}(0, \sigma(f, \beta)^2) \quad \text{as } n \rightarrow \infty,$$

and the asymptotic variance is given by

$$\sigma(f, \beta)^2 = \sigma(f)^2 + \int \pi(\text{d}x) \mathcal{Q}(x, \text{d}A) [\text{var}_{\kappa_{x,A}}(\beta_{x,A} - F) - \text{var}_{\kappa_{x,A}}(F)], \tag{3.6}$$

with  $\sigma(f)^2 = \langle \pi, F^2 - (PF)^2 \rangle$ .

- (iii) *The asymptotic variance  $\sigma(f, \beta)^2$  is minimal for  $\beta_{x,A} = F$  and*

$$\begin{aligned} \sigma(f, F)^2 &= \int \pi(\text{d}x) \left( \int \mathcal{Q}(x, \text{d}A) \langle \kappa_{x,A}, F \rangle^2 - \left( \int \mathcal{Q}(x, \text{d}A) \langle \kappa_{x,A}, F \rangle \right)^2 \right) \\ &\leq \sigma(f)^2. \end{aligned} \tag{3.7}$$

*Proof.* We shall prove the theorem when  $X_0$  is distributed according to  $\pi$ . The general case follows from Proposition 17.1.6 of [9], since  $X$  is Harris recurrent.

We set, for  $n \geq 1$ ,

$$\Delta M_n = F(X_n) - PF(X_{n-1}) + \eta(X_{n-1}, A_n, X_n),$$

where

$$\eta(x, A, y) = \sum_{\tilde{x} \in A} (\kappa(x, A, \tilde{x}) - \mathbf{1}_{\{y=\tilde{x}\}}) \beta(x, A, \tilde{x}).$$

Note that  $\Delta M_n$  is square integrable and that  $E[\Delta M_{n+1} \mid \mathcal{G}_n] = 0$ , where  $\mathcal{G}_n$  is the  $\sigma$ -field generated by  $X_0$  and  $(A_i, X_i)$  for  $1 \leq i \leq n$ . In particular,  $M = (M_n, n \geq 0)$  with  $M_n = \sum_{k=1}^n \Delta M_k$  is a martingale with respect to the filtration  $(\mathcal{G}_n, n \geq 0)$ . Using the fact that  $F$  solves the Poisson equation, we also have

$$\mathfrak{J}_n(f, \beta) = \frac{1}{n} M_n - \frac{1}{n} P F(X_n) + \frac{1}{n} P F(X_0) + \langle \pi, f \rangle. \tag{3.8}$$

As  $\langle \pi, F^2 \rangle < \infty$  implies that  $\langle \pi, |P F| \rangle < \infty$ , we deduce from Theorem 17.3.3 of [9] that, a.s.,  $\lim_{n \rightarrow \infty} (1/n) P F(X_n) = 0$ . In particular, part (i) will be proved as soon as we check that, a.s.,  $\lim_{n \rightarrow \infty} M_n/n = 0$ .

We easily compute the bracket of  $M_n$ :

$$\langle M \rangle_n = \sum_{k=1}^n E[\Delta M_k^2 \mid \mathcal{G}_{k-1}] = \sum_{k=1}^n h(X_{k-1}),$$

with

$$h(x) = P(F^2)(x) - (P F(x))^2 + \int \mathcal{Q}(x, dA) [-2 \text{cov}_{\kappa(x, A, \cdot)}(\beta_{x, A}, F) + \text{var}_{\kappa(x, A, \cdot)}(\beta_{x, A})].$$

Elementary computation yields

$$-2 \text{cov}_{\kappa(x, A, \cdot)}(\beta_{x, A}, F) + \text{var}_{\kappa(x, A, \cdot)}(\beta_{x, A}) = \text{var}_{\kappa(x, A, \cdot)}(\beta_{x, A} - F) - \text{var}_{\kappa(x, A, \cdot)}(F).$$

Since  $\langle \pi, F^2 \rangle < \infty$  and  $\int \pi(dx) \mathcal{Q}(x, dA) \kappa(x, A, dy) \beta(x, A, y)^2 < \infty$ ,  $h$  is  $\pi$  integrable. We set  $\sigma(f, \beta)^2 = \langle \pi, h \rangle$ , that is,  $\sigma(f, \beta)^2$  is given by (3.6), thanks to (2.5) and the fact that  $\pi$  is invariant for  $P$ . Theorem 17.3.2 of [9] asserts that, a.s.,  $\lim_{n \rightarrow \infty} \langle M \rangle_n/n = \langle \pi, h \rangle$ . Then Theorem 1.3.15 of [6] implies that, a.s.,  $\lim_{n \rightarrow \infty} M_n/n = 0$ . This completes the proof of part (i).

The proof of part (ii) relies on the central limit theorem for martingales; see Theorem 2.1.9 of [6]. We have already proved that, a.s.,  $\lim_{n \rightarrow \infty} \langle M \rangle_n/n = \sigma(f, \beta)^2$ . Let us now check Lindeberg’s condition. Note that theorem 17.3.2 of [9] implies that, for any  $a > 0$ , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[\Delta M_k^2 \mathbf{1}_{\{\Delta M_k^2 > a\}} \mid \mathcal{G}_{k-1}] = \langle \pi, h_a \rangle,$$

where  $h_a(x) = E[\Delta M_1^2 \mathbf{1}_{\{\Delta M_1^2 > a\}} \mid X_0 = x]$ . Note that  $0 \leq h_a \leq h$  and that  $(h_a, a > 0)$  decreases to 0 as  $a$  goes to  $\infty$ . We deduce that, a.s.,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n E[\Delta M_k^2 \mathbf{1}_{\{\Delta M_k^2 > \sqrt{n}\}} \mid \mathcal{G}_{k-1}] \leq \limsup_{a \rightarrow \infty} \langle \pi, h_a \rangle = 0.$$

This gives Lindeberg’s condition. We then deduce that  $(M_n/\sqrt{n}, n \geq 1)$  converges in distribution to  $\mathcal{N}(0, \sigma(f, \beta)^2)$ . Then we use (3.8) and the fact that, a.s.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} (P F(X_{n+1}))^2 = 0$$

(thanks to Theorem 17.3.3 of [9]) to complete the proof of part (ii).

The proof of part (iii) is as follows. The asymptotic variance  $\sigma(f, \beta)^2$  is minimal when  $\text{var}_{\kappa_{x,A}}(\beta_{x,A} - F) = 0$ , that is, at least for  $\beta_{x,A} = F$ . Of course,  $\sigma(f, F)^2 \leq \sigma(f, 0)^2 = \sigma(f)^2$ . Using (2.5), the fact that  $\pi$  is invariant for  $P$ , and definition (3.4) of  $P$ , we obtain

$$\begin{aligned} \sigma(f)^2 &= \langle \pi, PF^2 \rangle - \langle \pi, (PF)^2 \rangle \\ &= \int \pi(dx) \mathcal{Q}(x, dA) \langle \kappa_{x,A}, F^2 \rangle - \int \pi(dx) \left( \int \mathcal{Q}(x, dA) \langle \kappa_{x,A}, F \rangle \right)^2. \end{aligned}$$

The expression of  $\sigma(f, F)^2$  follows from (3.6).

#### 4. The Boltzmann case

We work in the general setting of Section 3 with the Boltzmann selection kernel  $\kappa$  given by (3.3) (or simply (2.14) when  $E$  is finite). The next proposition generalizes Proposition 2.5. It ensures that the asymptotic variance of the waste-recycling algorithm,  $\sigma(f, f)^2$ , is smaller than that of the standard Metropolis–Hastings algorithm,  $\sigma(f)^2$ , and that  $b \mapsto \sigma(f, bf)^2$  is minimal at  $b_\star$  given by (2.7). At the same time, we show that this  $\sigma(f)^2$  variance is at least divided by 2 for the optimal choice  $\beta(x, A, y) = F(y)$  in our control variate approach.

For  $f$  such that  $\langle \pi, f^2 \rangle < \infty$ , we set  $f_0 = f - \langle \pi, f \rangle$  and

$$\Delta(f) = \frac{1}{2} \int \pi(dx) P(x, dy) (f_0(x) + f_0(y))^2 = \langle \pi, f_0(f_0 + Pf_0) \rangle. \tag{4.1}$$

Note that the second equality in (4.1) is a consequence of the invariance of  $\pi$  with respect to  $P$ .

**Proposition 4.1.** *We assume that  $X$  is Harris recurrent, that  $\langle \pi, f^2 \rangle < \infty$ , and that there exists a solution  $F$  to the Poisson equation  $F - PF = f - \langle \pi, f \rangle$  such that  $\langle \pi, F^2 \rangle < \infty$ . We consider the Boltzmann case: the selection kernel  $\kappa$  is given by (3.3). For  $\beta(x, A, y)$  respectively equal to  $F(y)$  and  $f(y)$ , we have*

$$\sigma(f, F)^2 = \frac{1}{2}(\sigma(f)^2 - \text{var}_\pi(f)) \quad \text{and} \quad \sigma(f, f)^2 = \sigma(f)^2 - \Delta(f).$$

The nonnegative term  $\Delta(f)$  is positive when  $\text{var}_\pi(f) > 0$ .

Furthermore, if  $\text{var}_\pi(f) > 0$  then  $\langle \pi, f^2 - fPf \rangle = \frac{1}{2} E_\pi[(f(X_0) - f(X_1))^2]$  is positive, the function  $b \mapsto \sigma(f, bf)^2$  is minimal at

$$b_\star = \frac{\langle \pi, f^2 \rangle - \langle \pi, f \rangle^2}{\langle \pi, f^2 - fPf \rangle}, \tag{4.2}$$

and  $b_\star > 1$  when  $\sigma(f, f)^2 > 0$ .

*Proof.* Recall the notation from Example 3.1. We set  $\kappa_A^B(dy) = \kappa^B(A, dy)$ . For  $g$  and  $h$  real-valued functions defined on  $E$ , we have

$$\begin{aligned} \int \pi(dx) \mathcal{Q}(x, dA) \langle \kappa_A^B, g \rangle \langle \kappa_A^B, h \rangle &= \int N_0(dA) r_A(dx) \langle \kappa_A^B, g \rangle \langle \kappa_A^B, h \rangle \\ &= \int N_0(dA) \langle r_A, g \rangle \langle \kappa_A^B, h \rangle \\ &= \int \pi(dx) \mathcal{Q}(x, dA) g(x) \langle \kappa_A^B, h \rangle \\ &= \langle \pi, gPh \rangle, \end{aligned} \tag{4.3}$$

where we have used (3.3) for the second equality. Using this equality with  $h = g = F$  in the first term of the expression of  $\sigma(f, F)^2$  given in (3.7), we obtain

$$\sigma(f, F)^2 = \langle \pi, FPF - (PF)^2 \rangle = \frac{1}{2} \langle \pi, F^2 - (PF)^2 - (F - PF)^2 \rangle = \frac{1}{2} (\sigma(f)^2 - \text{var}_\pi(f)),$$

where we have used the Poisson equation (2.6) for the last equality.

We also obtain

$$\begin{aligned} & \int \pi(dx) \mathcal{Q}(x, dA) [\text{var}_{\kappa_A^B}(bf - F) - \text{var}_{\kappa_A^B}(F)] \\ &= \int \pi(dx) \mathcal{Q}(x, dA) [\langle \kappa_A^B, (bf - F)^2 \rangle - \langle \kappa_A^B, bf \rangle^2 + 2 \langle \kappa_A^B, bf \rangle \langle \kappa_A^B, F \rangle \\ & \quad - \langle \kappa_A^B, F \rangle^2 - \langle \kappa_A^B, F^2 \rangle + \langle \kappa_A^B, F \rangle^2] \\ &= \langle \pi, b^2 f^2 - 2bfF - b^2 fPf + 2fPF \rangle \\ &= b^2 \langle \pi, f^2 - fPf \rangle - 2b(\langle \pi, f^2 \rangle - \langle \pi, f \rangle^2), \end{aligned}$$

where we have used (4.3) for the second equation and (2.6) for the last equality. We deduce from (3.6) with  $\beta_{x,A} = bf$  that

$$\sigma(f, bf)^2 - \sigma(f)^2 = b^2 \langle \pi, f^2 - fPf \rangle - 2b(\langle \pi, f^2 \rangle - \langle \pi, f \rangle^2).$$

We first check that  $\text{var}_\pi(f) > 0$  implies that  $\langle \pi, f^2 - fPf \rangle > 0$ . If, when  $X_0$  is distributed according to  $\pi$ , a.s.,  $f(X_1) = f(X_0)$ , then, a.s.,  $k \mapsto f(X_k)$  is constant and, by the ergodic theorem, this constant is equal to  $\langle \pi, f \rangle$ . Therefore,  $\text{var}_\pi(f) > 0$  implies positivity of  $\langle \pi, f^2 - fPf \rangle$ , which is equal to  $\frac{1}{2} E_\pi[(f(X_0) - f(X_1))^2]$  by the reversibility of  $\pi$  with respect to  $P$ . Hence, when  $\text{var}_\pi(f) > 0$ , then  $b \mapsto \sigma(f, bf)^2$  is minimal for  $b = b_\star$  defined by (4.2).

For the choice  $b = 1$ , we obtain

$$-\sigma(f, f)^2 + \sigma(f)^2 = \langle \pi, f(f + Pf) \rangle - 2 \langle \pi, f \rangle^2 = \Delta(f) = \text{var}_\pi(f) + \langle \pi, f_0 P f_0 \rangle. \tag{4.4}$$

By (4.3),  $\langle \pi, f_0 P f_0 \rangle = \int \pi(dx) \mathcal{Q}(x, dA) \langle \kappa_A^B, f_0 \rangle^2 \geq 0$  and  $\Delta(f)$  is positive when  $\text{var}_\pi(f) > 0$ .

Moreover, the difference  $\langle \pi, fPf \rangle - \langle \pi, f \rangle^2 = \langle \pi, f_0 P f_0 \rangle$  is nonnegative thanks to (4.3) and when it is equal to 0, then (4.3) implies that  $\langle \pi, f_0 P g \rangle = \langle \pi, g P f_0 \rangle = 0$  for each function  $g$  on  $E$  such that  $\langle \pi, g^2 \rangle < +\infty$ . In this case, by (4.4),

$$\begin{aligned} \sigma(f, f)^2 &= \sigma(f)^2 + \sigma(f, f)^2 - \sigma(f)^2 \\ &= \langle \pi, (F + PF)(F - PF) \rangle - \text{var}_\pi(f) \\ &= \langle \pi, (f_0 + 2PF)f_0 \rangle - \text{var}_\pi(f) \\ &= 0. \end{aligned}$$

Hence, when  $\text{var}_\pi(f) > 0$  and  $\sigma(f, f)^2 > 0$ , then we have  $\langle \pi, f_0 P f_0 \rangle > 0$  and  $b_\star > 1$ .

### 5. Further results in the single-proposal case

The Metropolis–Hastings algorithm corresponds to the single-proposal case, which is the particular case of the multi-proposal algorithm of Section 3 where  $\mathcal{Q}(x, \cdot)$  gives full weight to the set of subsets of  $E$  (not assumed to be finite) containing  $x$  and at most one other element of  $E$ . The acceptance probability is then given by  $\rho(x, y) = \kappa(x, \{x, y\}, y)$  and the selection

kernel  $Q(x, \cdot)$  is the image of  $\mathcal{Q}(x, \cdot)$  by any measurable mapping such that the image of  $\{x, y\}$  is  $y$ . See Remark 2.3 in the particular case of  $E$  finite. Equation (3.1) is then equivalent to the following generalization of (2.1):

$$\pi(dx)Q(x, dy)\rho(x, y) = \pi(dy)Q(y, dx)\rho(y, x). \tag{5.1}$$

Moreover, the transition kernel of the Markov chain  $X$  is given by

$$\begin{aligned} \mathbf{1}_{\{y \neq x\}} P(x, dy) &= \mathbf{1}_{\{y \neq x\}} \rho(x, y)Q(x, dy) \\ \text{and } P(x, \{x\}) &= 1 - \int_{z \neq x} \rho(x, z)Q(x, dz). \end{aligned} \tag{5.2}$$

Motivated by the study of the WRMC algorithm which corresponds to  $\psi = f$  and of the optimal choice  $\psi = F$ , we first derive more convenient expressions for  $\sigma(f, \psi)^2$  in the single-proposal framework. We then use this new expression to construct a counterexample such that  $\sigma(f, f)^2 > \sigma(f)^2$ . Furthermore, when  $\rho(x, y) + \rho(y, x)$  is constant on  $E_*^2 = E^2 \setminus \{(x, x) : x \in E\}$ , again using the expression of  $\sigma(f, \psi)^2$ , we compute the value of  $b$  such that  $\sigma(f, bf)^2$  is minimal and check that  $\sigma(f, f)^2 < \sigma(f)^2$  as soon as  $f$  is nonconstant.

**5.1. Another expression of the asymptotic variance**

We recall that in the notation  $E_\pi$ , the subscript  $\pi$  means that  $X_0$  is distributed according to  $\pi$ .

**Lemma 5.1.** *We assume that  $\langle \pi, f^2 \rangle < \infty$  and that there exists a solution  $F$  to the Poisson equation (2.6) such that  $\langle \pi, F^2 \rangle < +\infty$ . Let  $\psi$  be square integrable:  $\langle \pi, \psi^2 \rangle < \infty$ . In the single-proposal case we have*

$$\begin{aligned} \sigma(f, \psi) &= \sigma(f)^2 - E_\pi[(1 - \rho(X_0, X_1))(F(X_1) - F(X_0))^2] \\ &\quad + E_\pi[(1 - \rho(X_0, X_1))(\psi(X_1) - F(X_1) - \psi(X_0) + F(X_0))^2]. \end{aligned}$$

*Proof.* In the single-proposal case,  $\kappa(x, \{x, y\}, y) = 1 - \kappa(x, \{x, y\}, x) = \rho(x, y)$  for  $x \neq y$ . Therefore, for a real-valued function  $g$  defined on  $E$ , we have

$$\text{var}_{\kappa(x, \{x, y\}, \cdot)}(g) = \rho(x, y)(1 - \rho(x, y))(g(y) - g(x))^2.$$

Thus, we deduce that

$$\begin{aligned} &\int_{E_*^2} \pi(dx)Q(x, dy) \text{var}_{\kappa(x, \{x, y\}, \cdot)}(g) \\ &= \int_{E_*^2} \pi(dx)Q(x, dy)\rho(x, y)(1 - \rho(x, y))(g(y) - g(x))^2 \\ &= \int_{E_*^2} \pi(dx)P(x, dy)(1 - \rho(x, y))(g(y) - g(x))^2 \\ &= E_\pi[(1 - \rho(X_0, X_1))(g(X_1) - g(X_0))^2], \end{aligned}$$

where we have used (5.2) for the second equality. Substituting this formula with  $g = \psi - F$  and  $g = F$  into (3.6) gives the result.

Taking  $\psi = F$  and  $\psi = f$  in the previous lemma gives the following corollary.



**Corollary 5.1.** *We assume that  $\langle \pi, f^2 \rangle < \infty$  and that there exists a solution  $F$  to the Poisson equation (2.6) such that  $\langle \pi, F^2 \rangle < +\infty$ . In the single-proposal case we have*

$$\begin{aligned} \sigma(f, F)^2 - \sigma(f)^2 &= -E_\pi[(1 - \rho(X_0, X_1))(F(X_1) - F(X_0))^2], \\ \sigma(f, f)^2 - \sigma(f)^2 &= -E_\pi[(1 - \rho(X_0, X_1)) \\ &\quad \times [(F(X_1) - F(X_0))^2 - (PF(X_1) - PF(X_0))^2]]. \end{aligned}$$

**5.2. A counterexample**

We construct a counterexample such that  $\sigma(f, f)^2 > \sigma(f)^2$  in the Metropolis case, thus proving the statements concerning this case in Proposition 2.3. This counterexample is also such that the optimal choice  $\psi = F$  does not achieve variance reduction:  $\sigma(f, F)^2 = \sigma(f)^2$ . Let  $P$  be an irreducible transition matrix on  $E = \{a, b, c\}$ , with invariant probability measure  $\pi$  such that  $P$  is reversible with respect to  $\pi$ ,

$$P(a, b) > 0, \quad P(a, a) > 0, \quad \text{and} \quad P(a, c) \neq P(b, c).$$

Let  $f$  be defined by  $f(x) = \mathbf{1}_{\{x=c\}} - P(x, c)$  for  $x \in E$ . We have

$$\langle \pi, f \rangle = \pi(c) - \sum_{x \in E} \pi(x)P(x, c) = 0.$$

The function  $F(x) = \mathbf{1}_{\{x=c\}}$  solves the Poisson equation (2.6):  $F - PF = f - \langle \pi, f \rangle$ .

Let  $\rho \in (P(a, b)/(P(a, a) + P(a, b)), 1)$ . We set

$$Q(x, y) = \begin{cases} \frac{P(a, b)}{\rho} & \text{if } (x, y) = (a, b), \\ P(a, a) - P(a, b)\left(\frac{1}{\rho} - 1\right) & \text{if } (x, y) = (a, a), \\ P(x, y) & \text{otherwise.} \end{cases}$$

We choose

$$\rho(x, y) = \begin{cases} \rho & \text{if } (x, y) = (a, b), \\ 1 & \text{otherwise.} \end{cases}$$

Since  $\rho(a, b)\pi(a)Q(a, b) = \rho\pi(a)P(a, b)/\rho$ , we have  $\rho(x, y)\pi(x)Q(x, y) = \pi(x)P(x, y)$  for all  $x \neq y \in E$ . Equation (2.1) follows from the reversibility of  $\pi$  for  $P$ . Note also that (2.2) holds with  $\gamma(u) = \min(1, u)$ .

By construction, the matrix  $P$  satisfies (2.3). By Corollary 5.1 we have  $\sigma(f, F)^2 - \sigma(f)^2 = 0$  and

$$\sigma(f, f)^2 - \sigma(f)^2 = \pi(a)P(a, b)(1 - \rho)(P(b, c) - P(a, c))^2 > 0. \tag{5.3}$$

Let us illustrate these results by simulation for the following specific choice:

$$\begin{aligned} \pi &= \frac{1}{10} \begin{pmatrix} 6 \\ 3 \\ 1 \end{pmatrix}, \quad P = \frac{1}{60} \begin{pmatrix} 38 & 21 & 1 \\ 42 & 0 & 18 \\ 6 & 54 & 0 \end{pmatrix}, \quad \rho = \frac{4}{10}, \\ \text{and } Q &= \frac{1}{120} \begin{pmatrix} 13 & 105 & 2 \\ 84 & 0 & 36 \\ 12 & 108 & 0 \end{pmatrix}. \end{aligned}$$

Then  $\sigma(f)^2 - \sigma(f, f)^2 = -0.010\ 115$  amounts to 14% of  $\sigma(f)^2 \simeq 0.072\ 833\ 3$ .

TABLE 1.

$n$	$\sigma_n^2$	$\sigma_{\text{WRMC},n}^2$	$\sigma_n^2 - \sigma_{\text{WRMC},n}^2$
1	[0.1213, 0.1339]	[0.1116, 0.1241]	[0.0091, 0.0104]
2	[0.0728, 0.0779]	[0.0758, 0.0815]	[-0.0041, -0.0025]
5	[0.0733, 0.0791]	[0.0798, 0.0859]	[-0.0075, -0.0058]
10	[0.0718, 0.0772]	[0.0800, 0.0859]	[-0.0094, -0.0074]
100	[0.0702, 0.0751]	[0.0803, 0.0858]	[-0.0114, -0.0092]
1000	[0.0719, 0.0769]	[0.0811, 0.0867]	[-0.0105, -0.0083]

Using  $N = 10\,000$  simulations, in Table 1 we give estimations of the variances  $\sigma_n^2$  of  $I_n(f)$  and  $\sigma_{\text{WRMC},n}^2$  of  $I_n(f, f)$ , and of the difference  $\sigma_n^2 - \sigma_{\text{WRMC},n}^2$  with asymptotic confidence intervals at level 95%. The initial variable  $X_0$  is generated according to the reversible probability measure  $\pi$ .

**5.3. Case of a constant  $\rho(x, y) + \rho(y, x)$**

Under the Boltzmann selection rule, according to Proposition 4.1, the asymptotic variance  $\sigma(f, f)^2$  of  $I_n^{\text{WRMC}}(f) = I_n(f, f)$  is smaller than the asymptotic variance  $\sigma(f)^2$  of  $I_n(f)$ , and  $\sigma(f, bf)$  is minimal for  $b = b_*$  given by (4.2). In the single-proposal case, the Boltzmann selection rule ensures that  $\rho(x, y) + \rho(y, x) = 1$  on  $E_*^2 = E^2 \setminus \{(x, x) : x \in E\}$ . It turns out that we are still able to prove the same results as soon as  $\rho(x, y) + \rho(y, x)$  is constant on  $E_*^2$ . Note that  $\text{var}_\pi(f) \geq 0$  and that the trivial case  $\text{var}_\pi(f) = 0$  corresponds to  $f$  constant  $\pi$ -a.s.

**Proposition 5.1.** *We assume that  $\langle \pi, f^2 \rangle < \infty$ , that  $\text{var}_\pi(f) > 0$ , and that there exists a solution  $F$  to the Poisson equation (2.6) such that  $\langle \pi, F^2 \rangle < \infty$ . We consider the single-proposal case and assume that there exists  $\alpha \in (0, 2)$  such that*

$$\pi(dx)Q(x, dy) \text{ a.s. on } E_*^2, \quad \rho(x, y) + \rho(y, x) = \alpha. \tag{5.4}$$

Then we have

(i)  $\langle \pi, f^2 - fPf \rangle = \frac{1}{2} E_\pi[(f(X_0) - f(X_1))^2]$  is positive,

(ii)

$$\begin{aligned} \sigma(f, \psi) - \sigma(f)^2 &= -\left(1 - \frac{\alpha}{2}\right) E_\pi[(F(X_1) - F(X_0))^2] \\ &\quad + \left(1 - \frac{\alpha}{2}\right) E_\pi[(\psi(X_1) - F(X_1) - \psi(X_0) + F(X_0))^2] \end{aligned} \tag{5.5}$$

for any real-valued function  $\psi$  on  $E$  such that  $\langle \pi, \psi^2 \rangle < \infty$ ,

(iii) the function  $b \mapsto \sigma(f, bf)^2$  is minimal at  $b_*$  given by (4.2) and  $b_* \geq 1/\alpha$ ,

(iv)  $\sigma(f, f)^2 - \sigma(f)^2 = -(2 - \alpha)\Delta(f) < 0$ , where  $\Delta(f)$  is given by (4.1).

*Proof.* Statement (i) follows from the proof of Proposition 4.1.

For statement (ii), note that, by the reversibility of  $\pi$ , we deduce from Lemma 5.1 that

$$\begin{aligned} \sigma(f, \psi) - \sigma(f)^2 &= -E_\pi[(1 - \rho(X_1, X_0))(F(X_1) - F(X_0))^2] \\ &\quad + E_\pi[(1 - \rho(X_1, X_0))(\psi(X_1) - F(X_1) - \psi(X_0) + F(X_0))^2]. \end{aligned}$$

This and Lemma 5.1 imply (5.5).

For statement (iii), using (5.5) with  $\psi = bf$ , it is straightforward to find that  $\sigma(f, bf)^2$  is minimal when  $b$  equals

$$\frac{E_\pi[(f(X_1) - f(X_0))(F(X_1) - F(X_0))]}{E_\pi[(f(X_1) - f(X_0))^2]} = \frac{\langle \pi, f(F - PF) \rangle}{\langle \pi, f^2 - fPf \rangle} = \frac{\langle \pi, f^2 \rangle - \langle \pi, f \rangle^2}{\langle \pi, f^2 - fPf \rangle} = b_\star.$$

Remarking that

$$b_\star = \frac{\langle \pi, f_0^2 \rangle}{\langle \pi, f_0^2 - f_0Pf_0 \rangle} = \frac{\langle \pi, f_0^2 \rangle}{\alpha \langle \pi, f_0^2 \rangle - \langle \pi, f_0Pf_0 + (\alpha - 1)f_0^2 \rangle}$$

and using Lemma 5.2, below, we deduce that  $b_\star \geq 1/\alpha$ .

We now prove statement (iv). Recall that  $f_0 = f - \langle \pi, f \rangle$ . Since

$$\langle \pi, f_0(f_0 + Pf_0) \rangle = (2 - \alpha) \text{var}_\pi(f) + \langle \pi, f_0Pf_0 + (\alpha - 1)f_0^2 \rangle,$$

we deduce from Lemma 5.2 that  $\Delta(f)$  given by (4.1) is positive. We have

$$\begin{aligned} & \frac{1}{2} E_\pi[(f(X_1) - F(X_1) - f(X_0) + F(X_0))^2 - (F(X_1) - F(X_0))^2] \\ &= \frac{1}{2} E_\pi[(f_0(X_1) - f_0(X_0))^2] - E_\pi[(f_0(X_1) - f_0(X_0))(F(X_1) - F(X_0))] \\ &= \langle \pi, f_0^2 - f_0Pf_0 \rangle - 2\langle \pi, f_0(F - PF) \rangle \\ &= -\langle \pi, f_0(f_0 + Pf_0) \rangle, \end{aligned}$$

where we have used the facts that  $\pi$  is invariant for  $P$  and  $P$  is reversible with respect to  $\pi$  for the second equality, and the fact that  $F$  solves (2.6) for the last equality. We conclude using (5.5) with  $\psi = f$ .

**Lemma 5.2.** *Let  $h$  be a real-valued function defined on  $E$  such that  $\langle \pi, h^2 \rangle < \infty$ . Under hypothesis (5.4), we have*

$$\langle \pi, hPh + (\alpha - 1)h^2 \rangle \geq 0.$$

*Proof.* Using (5.2) then (5.4), we obtain

$$\begin{aligned} \langle \pi, hPh + (\alpha - 1)h^2 \rangle &= \int_{E_*^2} \pi(dx) Q(x, dy) \rho(x, y) h(x)h(y) \\ &\quad + \int_E \pi(dx) \left( \alpha - \int_E \mathbf{1}_{y \neq x} Q(x, dy) \rho(x, y) \right) h^2(x) \\ &= \int_{E_*^2} \pi(dx) Q(x, dy) [\rho(x, y)h(x)h(y) + \rho(y, x)h^2(x)] \\ &\quad + \alpha \int_E \pi(dx) Q(x, \{x\}) h^2(x). \end{aligned}$$

To conclude, it is enough to check that the first term on the right-hand side is nonnegative.

Using (5.4) and (5.1) for the first equality, we obtain

$$\begin{aligned} &\alpha \int_{E_*^2} \pi(dx) \mathcal{Q}(x, dy) [\rho(x, y)h(x)h(y) + \rho(y, x)h^2(x)] \\ &= \int_{E_*^2} \pi(dx) \mathcal{Q}(x, dy) \rho(y, x) [\rho(x, y)h(x)h(y) + \rho(y, x)h^2(x)] \\ &\quad + \int_{E_*^2} \pi(dy) \mathcal{Q}(y, dx) \rho(y, x) [\rho(x, y)h(x)h(y) + \rho(y, x)h^2(x)] \\ &= \int_{E_*^2} \pi(dx) \mathcal{Q}(x, dy) [\rho(y, x)h(x) + \rho(x, y)h(y)]^2 \\ &\geq 0. \end{aligned}$$

### 6. Other remarks

We work in the general setting of Section 3.

#### 6.1. About the estimator $I_n(f + P\psi - \psi)$

Motivated by Remark 2.2 on the study of  $I_n(f + P\psi - \psi)$ , we compute the asymptotic variance  $\tilde{\sigma}(f, \beta)^2$  of

$$\begin{aligned} I_n(f) + \frac{1}{n} \sum_{k=0}^{n-1} &\left( \int \mathcal{Q}(X_k, dA) \kappa(X_k, A, d\tilde{x}) \beta(X_k, A, \tilde{x}) - \beta(X_k, A_{k+1}, X_{k+1}) \right) \\ &= I_n(f) + \frac{1}{n} \sum_{k=0}^{n-1} (\mathbb{E}[\beta(X_k, A_{k+1}, X_{k+1}) \mid X_k] - \beta(X_k, A_{k+1}, X_{k+1})). \end{aligned}$$

Following the proof of Theorem 3.1, we find that the above estimator of  $\langle \pi, f \rangle$  is under the hypotheses of Theorem 3.1, convergent, and asymptotically normal with asymptotic variance

$$\tilde{\sigma}(f, \beta)^2 = \sigma(f, \beta)^2 + \int \pi(dx) [\text{var}_{\mathcal{Q}(x, \cdot)}(\kappa\beta_x - \kappa F_x) - \text{var}_{\mathcal{Q}(x, \cdot)}(\kappa F_x)],$$

where  $\text{var}_{\mathcal{Q}(x, \cdot)}(\varphi) = \int \mathcal{Q}(x, dA) \varphi(A)^2 - (\int \mathcal{Q}(x, dA) \varphi(A))^2$ ,  $\kappa\beta_x(A) = \langle \kappa_{x,A}, \beta_{x,A} \rangle$ , and  $\kappa F_x(A) = \langle \kappa_{x,A}, F \rangle$ .

Note that the sign of  $\tilde{\sigma}(f, \beta)^2 - \sigma(f, \beta)^2$  depends on  $\beta$  (take  $\beta_{x,A} = F$  and  $\beta_{x,A} = -F$ ).

#### 6.2. Changing the selection kernel in $\mathcal{J}_n$

Let  $\kappa' \neq \kappa$  be such that (3.1) (or simply (2.10) if  $E$  is finite) still holds when  $\kappa$  is replaced by  $\kappa'$ , and let  $\mathcal{J}'_n(\psi)$  and  $\mathcal{J}'_n(\beta)$  be defined like  $\mathcal{J}_n(\psi)$  and  $\mathcal{J}_n(\beta)$  with the chain  $X$  unchanged but with  $\kappa(X_k, A_{k+1}, \tilde{x})$  replaced by  $\kappa'(X_k, A_{k+1}, \tilde{x})$  in (2.11) and (3.5). Thus, we have

$$\mathcal{J}'_n(\psi) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{\tilde{x} \in A_{k+1}} (\kappa'(X_k, A_{k+1}, \tilde{x}) - \mathbf{1}_{\{X_{k+1}=\tilde{x}\}}) \psi(\tilde{x}).$$

Note that in general  $\sum_{\tilde{x} \in A_{k+1}} \kappa'(X_k, A_{k+1}, \tilde{x}) \psi(\tilde{x}) \neq \mathbb{E}[\psi(X_{k+1}) \mid X_k, A_{k+1}]$ .

In the single-proposal case, Frenkel [8] suggested that  $\mathcal{J}'_n(f)$  can also be used as a control variate. In general, for a real-valued function  $\beta$  defined on  $E \times \mathcal{P} \times E$ , the almost-sure

limit of  $\mathcal{J}'_n(\beta)$  is different from zero, which means that the estimator  $I_n(f) + \mathcal{J}'_n(\beta)$  of  $\langle \pi, f \rangle$  is not convergent. However, when  $\beta(x, A, \cdot) = \psi(\cdot)$ , Lemma 6.1, below, ensures that the estimator  $I_n(f) + \mathcal{J}'_n(\psi)$  of  $\langle \pi, f \rangle$  is convergent. It is also easy to prove that this estimator is asymptotically normal and compute the asymptotic variance, but we have not been able to compare it with the asymptotic variance  $\sigma(f)^2$  of  $I_n(f)$ .

**Lemma 6.1.** *We assume that  $X$  is Harris recurrent, that  $\langle \pi, f^2 \rangle < \infty$ , that there exists a solution  $F$  to the Poisson equation  $F - PF = f - \langle \pi, f \rangle$  such that  $\langle \pi, F^2 \rangle < \infty$ , and that  $\psi$  is such that  $\langle \pi, \psi^2 \rangle < \infty$ . Under these assumptions, the estimator  $I_n(f) + \mathcal{J}'_n(\psi)$  of  $\langle \pi, f \rangle$  is consistent: a.s.,  $\lim_{n \rightarrow \infty} I_n(f) + \mathcal{J}'_n(\psi) = \langle \pi, f \rangle$ .*

*Proof.* We set

$$\Delta R_n = \int \kappa'(X_{n-1}, A_n, d\tilde{x})\psi(\tilde{x}) - \int \mathcal{Q}(X_{n-1}, dA)\kappa'(X_{n-1}, A, d\tilde{x})\psi(\tilde{x}).$$

Note that  $\Delta R_n$  is square integrable and that  $E[\Delta R_{n+1} | \mathcal{G}_n] = 0$ , where  $\mathcal{G}_n$  is the  $\sigma$ -field generated by  $X_0$  and  $(A_i, X_i)$  for  $1 \leq i \leq n$ . In particular,  $R = (R_n, n \geq 0)$  with  $R_n = \sum_{k=1}^n \Delta R_k$  is a martingale with respect to the filtration  $(\mathcal{G}_n, n \geq 0)$ . Note that

$$\begin{aligned} \mathcal{J}'_n(\psi) &= \frac{1}{n}R_n + I_n(\gamma) - \frac{1}{n} \int \mathcal{Q}(X_n, dA)\kappa'(X_n, A, d\tilde{x})\psi(\tilde{x}) \\ &\quad + \frac{1}{n} \int \mathcal{Q}(X_0, dA)\kappa'(X_0, A, d\tilde{x})\psi(\tilde{x}), \end{aligned}$$

where  $\gamma(x) = \int \mathcal{Q}(x, dA)\kappa'(x, A, d\tilde{x})\psi(\tilde{x}) - \psi(x)$ . Following the proof of Theorem 3.1, we easily find that, a.s.,  $\lim_{n \rightarrow \infty} (1/n)R_n = 0$  and that, a.s.,

$$\lim_{n \rightarrow \infty} \mathcal{J}'_n(\psi) = \lim_{n \rightarrow \infty} I_n(\gamma) = \langle \pi, \gamma \rangle.$$

Using (3.1) satisfied by  $\kappa'$  instead of  $\kappa$ , we find that  $\langle \pi, \gamma \rangle = 0$ . This completes the proof of the lemma.

### Acknowledgements

We warmly thank Manuel Athènes (CEA Saclay) for presenting the waste-recycling Monte Carlo algorithm to us and Randal Douc (CMAP École Polytechnique) for numerous fruitful discussions. We also thank the anonymous referees for their valuable comments.

### References

- [1] ANDERSEN, H. C. AND DIACONIS, P. (2007). Hit and run as a unifying device. *J. Soc. Fran. Stat.* **148**, 5–28.
- [2] ATCHADÉ, Y. F. AND PERRON, F. (2005). Improving on the independent Metropolis–Hastings algorithm. *Statistica Sinica* **15**, 3–18.
- [3] ATHÈNES, M. (2007). Web ensemble averages for retrieving relevant information from rejected Monte Carlo moves. *Europ. Phys. J. B* **58**, 83–95.
- [4] CEPERLEY, D., CHESTER, G. V. AND KALOS, M. H. (1977). Monte Carlo simulation of a many fermion study. *Phys. Rev. B* **16**, 3081–3099.
- [5] DOUC, R. AND ROBERT, C. P. (2009). A vanilla Rao–Blackwellisation of Metropolis–Hastings algorithms. Preprint. Available at <http://arXiv.org/0904.2144>.
- [6] DUFLO, M. (1997). *Random Iterative Models* (Appl. Math. **34**). Springer, Berlin.
- [7] FRENKEL, D. (2004). Speed-up of Monte Carlo simulations by sampling of rejected states. *Proc. Nat. Acad. Sci. USA* **101**, 17571–17575.

- [8] FRENKEL, D. (2006). Waste-recycling Monte Carlo. In *Computer Simulations in Condensed Matter: From Materials to Chemical Biology* (Lecture Notes Phys. **703**), Springer, Berlin, pp. 127–137.
- [9] MEYN, S. P. AND TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
- [10] MUNOS, R. (2006). Geometric variance reduction in Markov chains: application to value function and gradient estimation. *J. Mach. Learn. Res.* **7**, 413–427.
- [11] PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika* **60**, 607–612.
- [12] ROBERT, C. P. AND CASELLA, G. (1999). *Monte Carlo Statistical Methods*. Springer, New York.