# A NOTE ON THE USE OF METROPOLIS-HASTINGS KERNELS IN IMPORTANCE SAMPLING

PIERRE DEL MORAL , ARNAUD DOUCET , AND AJAY JASRA

**Abstract.** It is shown that it is possible to use Metropolis-Hastings (M-H) kernels in importance sampling (IS) algorithms which are of cost $O(N^2)$, $N$ being the number of simulated samples. It might not have been previously realized.

**1. Introduction.** Throughout this note, suppose one is interested in sampling from a probability measure $\pi$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$, where $\pi$ admits a positive density, also denoted $\pi$, w.r.t Lebesgue measure $\lambda^{\text{Leb}}$.

Suppose IS is used to simulate from $\pi$, with the following importance distribution:

$$q(dy) = \int_{\mathbb{R}} \eta(x) K(x, dy) \lambda^{\text{Leb}}(dx)$$

where, for $q(x, y)$ a conditional probability density w.r.t $\lambda^{\text{Leb}}$;

$$\begin{aligned} K(x, dy) &= \alpha(x, y) q(x, y) \lambda^{\text{Leb}}(dy) + \delta_x(dy) r(x) \\ r(x) &= 1 - \int_{\mathbb{R}} \alpha(x, y) q(x, y) \lambda^{\text{Leb}}(dy) \\ \alpha(x, y) &= 1 \wedge \frac{\pi(y) q(y, x)}{\pi(x) q(x, y)} \end{aligned}$$

is a M-H kernel of invariant distribution $\pi$ and $\eta(x)\lambda^{\text{Leb}}(dx)$ is a probability measure, with $\eta(x) > 0 \ \forall x \in \mathbb{R}^n$. Intuitively, this is a sensible scheme, as $K$ should help guide samples to 'good' parts of the state-space. Unfortunately, such a scheme is not possible as the associated importance weight:

$$\frac{d\pi}{dq}(y) = \frac{\pi(y)}{\int_{\mathbb{R}^n} \eta(x) \alpha(x, y) q(x, y) \lambda^{\text{Leb}}(dx) + r(y)\eta(y)} \tag{1.1}$$

can not be computed: the first integral appearing in the denominator and the rejection probability, $r(y)$, are not typically known.

Consider the scenario, instead where one samples from $\eta$, $N-$times, $\{X^{(i)}\}_{1 \leq i \leq N}$ and uses the importance distribution:

$$q^N(dy) = \frac{1}{N} \sum_{i=1}^{N} K(x^{(i)}, dy). \tag{1.2}$$

In this scenario, the importance weight is exactly

$$\frac{d\pi}{dq^N}(y) = \frac{\pi(y)}{\frac{1}{N} \sum_{i=1}^{N} \alpha(x^{(i)}, y) q(x^{(i)}, y)} \tag{1.3}$$

for $y \neq x^{(i)}$ and it is zero otherwise.

A rigourous proof of this can be established along the lines of [3]. An informal proof follows from the fact that the probability of generating a candidate from $q^N(dy)$ such that $y \neq x^{(i)}$ is given by

$$\frac{1}{N} \sum_{i=1}^{N} \left(1 - r(x^{(i)})\right)$$

and the distribution of such a candidate is

$$\frac{\sum_{i=1}^{N}\alpha(x^{(i)},y)q(x^{(i)},y)\lambda^{\mathrm{Leb}}(dy)}{\sum_{i=1}^{N}\left(1-r(x^{(i)})\right)}.$$

Hence if one has the computational power to perform an $O(N^2)$ algorithm it is possible to use M-H kernels in an IS algorithm.

**2. Implications of the Result.** The main motivation of this note relates to our earlier paper [2]. In that paper, we state correctly that $q^N(dy)$ is not known exactly as the rejection probabilities $r(x^{(i)})$ are unknown. However, it was implicitly implied that it prevents one from using M-H kernels in proposals such as in (1.2). Our reasoning was as above; the theoretical Radon-Nikodym derivative given in (1.1) is unknown and requires the knowledge of the rejection probability $r(y)$. However, clearly, the correct Radon-Nikodym derivative is as (1.3) and does not require knowledge of the rejection probabilities $r(x^{(i)})$. This result might be useful in the context of population Monte Carlo algorithms (see e.g. [1] and the references therein). Such algorithms use 'standard' proposal mechanisms in order to simulate the samples, along with adaptation techniques for the proposals. This note shows that such methodology can use M-H kernels. This enriches the potential applications of this methodology.

## REFERENCES


[1] Beaumont, M. A., Cornuet, J. M., Marin, J. M. & Robert, C. P. (2009). Adaptivity for approximate Bayesian computation algorithms: a population Monte Carlo approach, *Biometrika*, (to appear).

[2] Del Moral, P., Doucet, A. & Jasra, A. (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B*, **68**, 411–436.

[3] Gottardo, R. & Raftery, A. E. (2008). Markov chain Monte Carlo for mixture of mutually singular distributions, *J. Comp. Graph. Stat.*, **17**, 949-975.