Journal of Computational and Applied Mathematics 319 (2017) 440–459

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Adaptive importance sampling Monte Carlo simulation for general multivariate probability laws

Reiichiro Kawai *

School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 17 January 2017

MSC:
65C05
93E35
60F05
60E05

Keywords:
Bypass distribution
Central limit theorem
Exponential family
Stochastic approximation
Variance reduction

ABSTRACT

We establish a parametric adaptive importance sampling variance reduction method for general multivariate probability laws. Employing the principle of bypass distributions makes it possible to develop adaptive algorithms without relying on particular properties of the target and proposal laws, both of which in the proposed framework are as general as the uniform law on the unit hypercube, without changing the sampling distribution at each iteration. We establish the asymptotic normality of the estimator of the desired mean and of the importance sampling parameter as the number of observations tends to infinity. Although implementation of the proposed methodology requires a small amount of initial work, it has the potential to yield substantial improvements in estimator efficiency in various general problem settings. To illustrate the applicability and effectiveness, we provide numerical results throughout, in which we apply exponential and normal bypass distributions, as well as demonstrate that well-known adaptive importance sampling formulations in the literature can be easily rewritten in the proposed framework.

1. Introduction

With significant increases in computational demands of Monte Carlo simulation, methods for increasing efficiency have attracted substantial interest, in particular for routines to be run frequently. It is thus worthwhile to carry out some initial analysis yielding more efficient simulation. In this spirit, in standard parametrized variance reduction methods, the parameter value is chosen in advance in the way to reduce, hopefully minimize, the estimator variance with differences in computing costs put into consideration. This parameter choice can be done, for instance, by taking a pure guess, or a more plausible guess from the past experience, or by running a pilot simulation for a while, etc., all without strong confidence of optimality. In particular, running a sufficiently long pilot simulation is a little contradictory from a practical point of view, in the sense that we might have been better off spending this effort running the crude Monte Carlo simulation longer, as well as, in the first place, the sufficient length of the pilot run is not obvious in advance.

Adaptive Monte Carlo variance reduction methods aim at concurrently running a Monte Carlo simulation and a search algorithm for optimal variance reduction parameters, where common replications can be used for both procedures without changing the sampling distribution at each iteration, rather than generating two independent sets of replications. Adaptive Monte Carlo methods require a certain amount of such initial work for its implementation, while it has the potential to provide significant variance reduction as a result. Equally importantly, the adaptive framework avoids the need for frequent

* Fax: +61 (0) 2 9351 4534.
E-mail address: reiichiro.kawai@sydney.edu.au.

http://dx.doi.org/10.1016/j.cam.2017.01.029
0377-0427/

recalibration of the parameters of the variance reduction techniques when changes occur in the experimental conditions governing system performance. The idea of adaptive Monte Carlo simulation and its practical use has been studied for a long time in various formulations and problem settings; for example, parametrized distributions [1–6], mixture densities [7–11], and (combinations of) control variates, importance sampling and stratified sampling [12–17], to mention just a few.

This paper is concerned with the construction and analysis of adaptive importance sampling variance reduction methods, for general multivariate probability laws, more precisely, the uniform law on the unit hypercube. Focusing on the uniform law is not a restriction but a generalization, in the sense that the expected value of a functional of a multivariate random vector can be rewritten with the standard uniform random vector in the same dimension with a suitable change of variables or the principle of inverse transform sampling. In the proposed framework, both target and proposal laws are simply the same uniform law on the unit hypercube. In particular, one class of parametric adaptive importance sampling methods is designed to adaptively reform the proposal law to make it closer to the target law (e.g. [10,17]), while the other class relies largely on special properties of the target distribution. For instance, a Gaussian random vector after an exponential change of measure is identical in law to the original Gaussian random vector with a suitable mean under the original probability measure (e.g. [1]), or a gamma random variable after an exponential change of measure is identical in law to a scaled original gamma random variable under the original probability measure (e.g. [14]).

The proposed framework is built upon a principle different from all those: it relies on no particular properties of the target and proposal laws. The key principle can be summarized as follows; first rewrite the expected value on the uniform law by a suitable parametric probability law with the principle of inverse transform sampling, secondly change the probability measure on the parametric law, thirdly rewrite the parametric law back to the uniform law under the original probability measure again with the principle of inverse transform sampling, and finally find optimal parameters introduced in the second step in the way to reduce the estimator variance. As is clear, the parametric law is employed in the second step solely to inject a parametrization into the expression of the expected value, whereas it does neither appear in the expected value in the final form nor change the underlying probability measure. For those reasons, we give the name of the *bypass* distribution to this parametric law in the second step. As a Monte Carlo simulation and a search algorithm for optimal variance reduction parameters are both expressed with common random elements under the original probability measure, the existing optimal parameter search techniques can be applied, such as the stochastic approximation [1,13–15,3,4,6] as well as the sample average approximation [2,18], under suitable technical conditions.

The rest of this paper is organized as follows. In Section 2, we begin with general notation and then briefly summarize background material on adaptive Monte Carlo variance reduction methods. In Section 3, we introduce the principle of bypass distributions so as to induce importance sampling. The choice of bypass distributions is quite flexible as far as technical conditions are satisfied, whereas theoretical over-complications often do not contribute to the practical effectiveness. In Section 4, we thus pay particular attention to the continuous canonical exponential family, a relatively simple yet wide class of continuous distributions. The derivation of the results entails rather lengthy proofs of somewhat routine nature. To avoid overloading the paper, we omit non-essential details in some instances. We provide numerical examples to demonstrate in Section 5 the procedure of the proposed method, as well as to illustrate the performance relative to the choice of bypass distributions. We also show that well-known adaptive importance sampling formulations in the literature can be easily rewritten in the proposed framework. In Section 6, we formulate the adaptive Monte Carlo simulation, along with convergence results and numerical illustrations of searching the parameter by the stochastic approximation and the sample average approximation. Finally, Section 7 concludes this study and highlights future research directions.

2. Problem setup

We begin with general notation which will be used throughout the paper. We use the notation $\mathbb{N} := \{1, 2, \ldots\}$ and denote by $|\cdot|$ and $\|\cdot\|$, respectively, the magnitude and the Euclidean norm. As usual, for a square matrix A, we denote by $|A|$, $\|A\|$, $A^\top$, $A^{\otimes 2}$ and $\sqrt{A}$, respectively, the determinant, a suitable matrix norm, the transpose, the outer product and a lower triangular matrix of the Cholesky decomposition of A, provided that those are well defined. We denote by ϕ, Φ and Φ^{-1}, respectively, the standard normal density function, the standard normal cumulative distribution function and its inverse. We denote by $\mathrm{Leb}(D)$, $\mathrm{int}(D)$, ∂D, $\overline{D}$ and $\mathscr{B}(D)$, respectively, the Lebesgue area, the interior, the boundary, the closure and the Borel σ-field of a domain D. We denote by $\mathbb{1}_D(\mathbf{x})$ the indicator function of a set D at $\mathbf{x}$. We let $\overset{\mathscr{L}}{=}$ and $\overset{\mathscr{L}}{\to}$ denote the identity and convergence in law. For the sake of simplicity, we use the notation ∂_x^q for the qth partial derivative with respect to the univariate variable x, as well as $\nabla_{\mathbf{x}}$ and $\mathrm{Hess}_{\mathbf{x}}$ indicate the gradient and the Hessian matrix with respect to the multivariate variable $\mathbf{x}$.

Throughout this paper, we are interested in constructing a fairly general framework of parametric adaptive importance sampling Monte Carlo methods for the integral

$$C := \int_{(0,1)^d} \Psi(\mathbf{u})\, d\mathbf{u} = \mathbb{E}_{\mathbb{P}}[\Psi(U)], \tag{2.1}$$

where Ψ is a function mapping from $(0, 1)^d$ to $\mathbb{R}$, and where U is a uniform random variables on $(0, 1)^d$ under the probability measure $\mathbb{P}$. We reserve the capital "C" for this integral value throughout the paper. To avoid triviality, we impose a finite