Journal of Computational Physics 314 (2016) 661–681

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Approximation of probability density functions by the Multilevel Monte Carlo Maximum Entropy method

Claudio Bierig, Alexey Chernov *

Carl von Ossietzky University, 26111 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 16 September 2015
Received in revised form 11 March 2016
Accepted 12 March 2016
Available online 21 March 2016

Keywords:
Multilevel Monte Carlo method
Maximum Entropy method
Kullback–Leibler divergence
Statistical moments
Moment matching

ABSTRACT

We develop a complete convergence theory for the Maximum Entropy method based on moment matching for a sequence of approximate statistical moments estimated by the Multilevel Monte Carlo method. Under appropriate regularity assumptions on the target probability density function, the proposed method is superior to the Maximum Entropy method with moments estimated by the Monte Carlo method. New theoretical results are illustrated in numerical examples.

1. Introduction

The Multilevel Monte Carlo Method (MLMC) is a recently established technique for efficient computation of an observable's statistics by approximate sampling in the case when generation of samples of different accuracy is possible. The method is particularly advantageous for complex problems with low regularity, typically resulting in high memory and CPU time demands. The idea is based on the observation that coarse sample approximations can be used as control variates for more accurate sample approximations and thereby reduce the variance of the Monte Carlo estimator. This family of methods has been introduced by M. Giles [1] for Itô stochastic differential equations arising in mathematical finance after similar ideas have been published in the earlier work by S. Heinrichs [2] on numerical quadrature. Since then MLMC has been extended to elliptic PDEs [3,4], parabolic problems [5], conservation laws [6], variational inequalities [7,8], multiscale PDEs [9], Kalman filtering [10] and other fields. The recent work [11] contains a recipe for an efficient evaluation of central statistical moments of arbitrary order. The aim of the present article is the further extension of the MLMC methodology for estimation of probability density functions.

Setting up probability density functions (PDF) on the basis of incomplete information on the observable is a prominent problem in statistics and information theory. One way to solve it is to recover the PDF from a truncated sequence of statistical moments (see the recent work by Giles et al. [12] for an alternative approach). This task (also known as solving the *truncated moment problem*) is by no means trivial and has been extensively studied in measure and probability theory [13–16]. It is well known that depending on the prescribed moments, the truncated moment problem may have no solution or multiple (infinitely many) solutions. The latter is typically the case when the truncated sequence of moments is admissible, i.e. it corresponds to some PDF (ruling out the case of negative even-order monomial moments and similar

* Corresponding author.
E-mail addresses: claudio.bierig@uni-oldenburg.de (C. Bierig), alexey.chernov@uni-oldenburg.de (A. Chernov).

http://dx.doi.org/10.1016/j.jcp.2016.03.027
0021-9991/

incompatibilities). However, in the presence of significant statistical and approximation errors, the sequence of *estimated* moments may become inadmissible even when the sequence of *exact* moments is admissible.

Assuming that the truncated moment sequence is admissible, one needs a criterion to select a PDF which is the "most appropriate" among infinitely many solutions to the truncated moment problem. The strategy of selecting the least biased estimate brings us to the concept of the Maximum Entropy (ME) method [17]. The ME solution is the (nonnegative) maximiser of the Shannon entropy constraint at the prescribed moment values. Obviously, the error of this approximation depends on the number of statistical moments and the accuracy of the *estimated* moments. Under appropriate assumptions the original constraint Maximum Entropy formulation is equivalent to the matching of moments with a density function whose logarithm is approximated by a polynomial. References [18,19] contain a rigorous error analysis of this class of ME methods, [18] also combines it with the Monte Carlo approach. The purpose of this work is to combine the Maximum Entropy approach with the Multilevel Monte Carlo estimation of moments and develop a rigorous error analysis in terms of i) the number of statistical moments, ii) statistical error and iii) discretization error. We derive complexity estimates for the proposed approach, test its performance on a set of synthetic problems with known PDFs, and demonstrate its applicability in a more realistic context: on a problem modelling contact of an elastic membrane with a rough random obstacle.

The outline of the paper is as follows. After a brief introduction to the Multilevel Monte Carlo and the Maximum Entropy methods in Section 2 we give a complete *a priori* error analysis for the proposed method in Section 3. In particular, we consider three different approximation methods for the set of the statistical moments: the Monte Carlo method based on exact sampling, the Monte Carlo method based on approximate sampling, and the Multilevel Monte Carlo approach. The error estimates naturally depend on the number of statistical moments, the sample size and the level of accuracy for approximate samples. In Section 4 we identify the optimal relation between these parameters and derive error-versus-cost relations for the three aforementioned methods. In Section 5 we give a series of numerical experiments illustrating convergence of the suggested Maximum Entropy approximations and compare them with a variant of Kernel Density Estimators available from the literature.

In the following, $\ln(\cdot)$ stands for the natural logarithm. We use a convention that for two scalar quantities f and g the notation $f \lesssim g$ means that there exists a nonnegative constant C independent of the approximation parameters such that $f \leq Cg$. The notation $f \sim g$ is equivalent to $f \lesssim g$ and $g \lesssim f$.

2. Preliminaries

In this section we recall some preliminary information needed for the subsequent analysis, see e.g. [20] and the references therein for the general framework of the multilevel Monte Carlo method (we utilise the notations from [8,11]), and [18,19,21] for the description of the Maximum Entropy method.

2.1. Multilevel Monte Carlo method

Suppose $(\Omega, \Sigma, \mathbb{P})$ is a probability space and X is a real-valued random variable which is not available for direct sampling. Instead, there exists an approximation X_ℓ to X, so that samples X_ℓ^i of X_ℓ can be generated. In this case the mean $\mathbb{E}[X]$ can be approximated by the sample average $E_M[X_\ell] := \frac{1}{M}\sum_{i=1}^M X_\ell^i$ of iid samples X_ℓ^i admitting the decomposition of the mean square error (MSE)

$$\|E_M[X_\ell] - \mathbb{E}[X]\|_{L^2}^2 = |\mathbb{E}[X - X_\ell]|^2 + \frac{1}{M}\operatorname{Var}[X_\ell] \tag{1}$$

where $\operatorname{Var}[X_\ell]$ is the variance of X_ℓ. The idea of the two-level Monte Carlo approach is to use samples from a coarser approximation $X_{\ell-1}$ to reduce the variance of the estimator. Indeed, for the two-level estimator it holds that

$$\begin{aligned}\|E_{M_\ell}[X_\ell - X_{\ell-1}] + E_{M_{\ell-1}}[X_{\ell-1}] - \mathbb{E}[X]\|_{L^2}^2 &= |\mathbb{E}[X - X_\ell]|^2 \\ &+ \frac{1}{M_\ell}\operatorname{Var}[X_\ell - X_{\ell-1}] + \frac{1}{M_{\ell-1}}\operatorname{Var}[X_{\ell-1}]\end{aligned}$$

where $E_{M_\ell}[X_\ell - X_{\ell-1}]$ and $E_{M_{\ell-1}}[X_{\ell-1}]$ are based on independent samples.

This situation occurs for example when X depends on a solution of an ODE or a PDE which is not available in closed form, but can be computed approximately, e.g. by the Finite Element Method or another numerical approximation method. In this setting the parameter ℓ plays the role of a discretization parameter. It is plausible that the samples of the fine approximation X_ℓ are better approximations to samples of X, but are typically more expensive to compute than samples of the coarse approximation $X_{\ell-1}$. The Multilevel Monte Carlo Method extends the two-level approach to multiple levels. In particular, the multilevel sample mean estimator is defined as

$$E^{\mathrm{ML}}[X] := \sum_{\ell=1}^{L} E_{M_\ell}[X_\ell - X_{\ell-1}], \quad X_0 := 0.$$