Monte Carlo method and sensitivity estimations

A. de Lataillade\textsuperscript{a,*}, S. Blanco\textsuperscript{b}, Y. Clergent\textsuperscript{b}, J.L. Dufresne\textsuperscript{c}, M. El Hafi\textsuperscript{a}, R. Fournier\textsuperscript{b}

\textsuperscript{a}Centre Energétique-Environnement, UMR CNRS 2392 Ecole des Mines, d’Albi-Carmaux, 31 allées des Sciences, 81000 Albi CT Cedex 09, France
\textsuperscript{b}Laboratoire d’Energétique UFR-PCA, Université Paul Sabatier, 31062 Toulouse Cedex, France
\textsuperscript{c}Laboratoire de Météorologie Dynamique, UMR CNRS 8359, Université de Paris VI, 75252 Paris Cedex 05, France

Received 30 November 2001; accepted 25 February 2002

Abstract

It is shown that, starting from any existing Monte Carlo algorithm for estimation of a physical quantity $A$, it is possible to implement a simple additional procedure that simultaneously estimates the sensitivity of $A$ to any problem parameter. The corresponding supplementary cost is very low as no additional random sampling is required. The principle is presented on a formal basis and simple radiative transfer examples are used for illustration. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Monte Carlo; Sensitivity analysis

1. Introduction

Monte Carlo methods are commonly used for the simulation of various transport phenomena [1–3]. The term Monte Carlo is, however, used to cover such a wide range of distinct statistical approaches that it is always a source of ambiguity to address any Monte Carlo methodology on a general basis. In the present paper, we restrict the analysis to methods that may be shown to rely on the generic Monte Carlo algorithm for the estimation of multiple integrals.

We address the question of estimating parametric sensitivities with Monte Carlo algorithms. Parametric sensitivity estimates are required in numerous physical contexts [4,5], in particular that of inverse radiative transfer problems as a basis for numerous retrieval algorithms [6,7]. For our purposes, sensitivities are mainly considered as a way of deriving first-order radiative transfer models for unstationary coupling with other complex phenomena such as natural convection [8] or chemical reactions in combustion systems [9]. In such contexts, the Monte Carlo method is of particular

\* Corresponding author. Present address: CERFACS, 42 Avenue Gaspard Coriolis, 31057 Toulouse Cedex 01, France. E-mail address: latailla@cerfacs.fr (A. de Lataillade).

0022-4073/02/$-see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0022-4073(02)00027-4
interest because of its ability to deal with complex geometries and/or complex spectral properties [10,11]. To our knowledge, the question of computing corresponding parametric sensitivities has not yet been addressed.

We first show that low-cost sensitivity computations can be easily associated with Monte Carlo methods, provided that the underlying integral formulation is explicit. This is seldom the case, however, for good reason, as we illustrate with a simple radiative transfer problem. The central aim of the paper is then to show how this problem may be effectively bypassed, making possible the computation of non-trivial sensitivity examples.

2. Principle (1)

Let us briefly recall the Monte Carlo algorithm in order to define the vocabulary and notation. If \( A \) is the integral (to be computed) of a function \( f \) over a multiple dimension domain \( \mathcal{D} \), the first step is to introduce arbitrarily a non-zero probability density function \( p \) over \( \mathcal{D} \) and to formulate \( A \) as the expectation of \( W = f(X)/p(X) \) where \( X \) is a random variable distributed according to \( p \):

\[
A = \int_{\mathcal{D}} f(x) \, dx = \int_{\mathcal{D}} \frac{f(x)}{p(x)} \, p(x) \, dx = E \left( \frac{f(X)}{p(X)} \right) = E(W).
\]  

(1)

The second step is to sample numerically a large number, \( N \), of realizations \( x_1, x_2, \ldots, x_N \) of the random variable \( X \), that is to say according to the arbitrarily chosen probability density function \( p \). Then, for each realization \( x_i \), the so-called Monte Carlo weight \( w_i = f(x_i)/p(x_i) \) is computed and the corresponding average value is retained for numerical estimation of \( A \) (see Appendix):

\[
A \approx \frac{1}{N} \sum_{i=1}^{N} w_i.
\]  

(2)

Once presented this way, the question of implementing sensitivity estimations inside Monte Carlo algorithms is of no particular subtlety. If the function \( f \) to be integrated now depends on any parameter \( \lambda \), the sensitivity of \( A \) to \( \lambda \), noted \( \partial \lambda A \) hereafter, may be expressed as

\[
\partial \lambda A = \int_{\mathcal{D}} \partial \lambda f(x; \lambda) \, dx = \int_{\mathcal{D}} \frac{\partial \lambda f(x; \lambda)}{p(x)} \, p(x) \, dx = E \left( \frac{\partial \lambda f(X; \lambda)}{p(X)} \right) = E(\partial \lambda W)
\]  

(3)

and then estimated as

\[
\partial \lambda A \approx \frac{1}{N} \sum_{i=1}^{N} \partial \lambda w_i.
\]  

(4)

This means that for any Monte Carlo algorithm inside the above-described family, where Monte Carlo weights are computed analytically on the basis of randomly generated variables, sensitivity estimations may be performed by computing the weight sensitivities and taking the average. This operation generally represents a very low additional cost because no further sampling is required (the same set of random generations is used for sensitivity estimations as for integral estimation). Fig. 1 proposes a schematic representation of this simple way of associating sensitivity estimations to most standard Monte Carlo algorithms.
Fig. 1. Schematic representation of a Monte Carlo algorithm for estimation of an integral $A$ together with its sensitivity to a parameter $\lambda$.

The first conceptual difficulty is that for most Monte Carlo algorithms the underlying integral formulation is not explicit. The reason is to be found in what is commonly considered as the main quality of Monte Carlo algorithms: they preserve strong analogies with the involved physics. This means that a Monte Carlo algorithm may be successfully designed and implemented by reproducing physical statistical events one after another, without requiring any complete mathematical translation. We consider below a simple radiative transfer example for illustration.

3. Example (1)

Consider the surface radiative power emitted by a blackbody and absorbed by an adjacent plane parallel to uniform slab of gray absorbing material of thickness $z$ (see Fig. 2). The blackbody temperature is denoted by $T$ and the slab absorption coefficient by $k$. The addressed problem is hereafter to compute the sensitivity to $k$ of this absorbed power.

Classically, the design of a corresponding Monte Carlo algorithm would strictly follow available statistical models of photon emission and photon transport. For instance, the surface flux emitted by the blackbody, $\sigma T^4$, is split into $N$ bundles of power $\sigma T^4/N$ and for each bundle an emission
direction \( \bar{u} \) is randomly generated according to the well-known distribution of blackbody-emission directions (Lambert law): 
\[
p_{\Omega,e}(\bar{u}) = \frac{\bar{u} \bar{n}}{4\pi}
\]
where the vector \( \bar{n} \) is normal to the slab. Then the bundle is followed along a straight line through the absorbing layer, with an exponential extinction to represent absorption, and when leaving the layer the remaining bundle-power is 
\[
\frac{\sigma T^4}{N} \exp\left(-\frac{kz}{\cos(\theta)}\right)
\]
where \( \theta \) is the angle between the emission direction and the normal. 
\[
(\sigma T^4/N)[1 - \exp(-kz/cos(\theta))]
\]
is therefore the bundle-power that was absorbed by the slab.

From these random angular generations, the absorbed flux is simply estimated as the sum of all the absorbed fractions:
\[
A \approx \sum_{i=1}^{N} \frac{\sigma T^4}{N} \left[ 1 - \exp\left(-\frac{kz}{\cos(\theta)}\right) \right].
\]
(5)

No formal integral expression is required to produce this algorithm. It is, however, very easy to identify the Monte Carlo weight as 
\[
w(\theta) = \frac{\sigma T^4}{N}[1 - \exp(-kz/cos(\theta))]
\]
and the sensitivity of \( A \) to the parameter \( k \) may therefore be estimated as
\[
\hat{c}_kA \approx \frac{1}{N} \sum_{i=1}^{N} \hat{c}_k w_i \quad \text{with} \quad \hat{c}_k w_i = \frac{z\sigma T^4}{\cos(\theta_i)} \exp\left(-\frac{kz}{\cos(\theta_i)}\right).
\]
(6)

4. Principle (2)

From this example, it could be concluded that sensitivity estimations within existing Monte Carlo algorithms are trivial to implement even if the formal integration is not explicit: (1) identifying the Monte Carlo weight expression, and (2) deriving it as a function of the considered parameter and taking the average. This is true only if the probability density function used for random generations is independent of the parameter (note that in Eq. (3) the probability density function \( p(x) \) has no parametric dependence on \( \lambda \)). In our first example, this was the case as the directional probability density function was independent of the absorption coefficient.

In the general case, the probability density function \( p \) chosen for random sampling of \( x_1, x_2, \ldots, x_N \) may depend on the considered parameter \( \lambda \) (see Fig. 1). Let us therefore consider that a given Monte Carlo algorithm is available, without the corresponding integral (which can be cumbersome
for multiple dimension problems as will be illustrated below), and that sensitivity estimations are required. All we know is that the addressed quantity \( A \) is estimated as the average of a Monte Carlo weight \( W \) that is a function of a random variable \( X \). The probability density function has a parametric dependence on \( /NAK \). We may formally write

\[
A = \int_\mathcal{D} w(x; \lambda) p(x; \lambda) \, dx
\]  

(7)

and the sensitivity of \( A \) to \( /NAK \) is

\[
\partial /NAK A = \int_\mathcal{D} \left[ \partial /NAK w(x; \lambda) p(x; \lambda) + w(x; \lambda) \partial /NAK p(x; \lambda) \right] \, dx
\]

\[
= \int_\mathcal{D} \left[ \partial /NAK w(x; \lambda) + w(x; \lambda) \frac{\partial /NAK p(x; \lambda)}{p(x; \lambda)} \right] p(x; \lambda) \, dx
\]

\[
= E \left( \partial /NAK W + W \frac{\partial /NAK p(X; \lambda)}{p(X; \lambda)} \right).
\]

(8)

It appears that a correction term is required in comparison with the simple case of an insensitive probability density function

\[
\partial /NAK A \approx \frac{1}{N} \sum_{i=1}^{N} \left[ \partial /NAK w_i + w_i \frac{\partial /NAK p(x_i; \lambda)}{p(x_i; \lambda)} \right].
\]

(9)

Implementing a sensitivity estimation inside any existing Monte Carlo algorithm is therefore a source of no practical difficulty, as all the information required for Eq. (9) is accessible by simple analysis of the considered original algorithm.

5. Example (2)

To illustrate such a practical implementation, let us reconsider the preceding example, the layer being now both absorbing (absorption coefficient \( k_a \)) and scattering (scattering coefficient \( k_s \)). An example of the application of standard Monte Carlo algorithms consists of emitting bundles at the source and following them until they exit the layer by a random walk corresponding to the pure diffusion assumption [1]. The traveled length before exit is then used to compute the extinction of the bundle-power according to the pure-absorption exponential extinction. As before, the algorithm starts with the random sampling of an emission direction \( \tilde{u}_1 \) for each of the \( N \) bundles of initial power \( \sigma T^4 / N \). Then a scattering path length \( \sigma_1 \) is sampled according to an exponential distribution on \( ]0, +\infty[ : p_\Sigma(\sigma_1; k_s) = k_s \exp(-k_s \sigma_1). This path length is used, starting at the emission point in the direction \( \tilde{u}_1 \), to define a first scattering location \( P_1 \) (see Fig. 3). If \( P_1 \) is outside the layer, the length \( l \) to the layer exit is used to compute the fraction of the bundle-power that is absorbed within the slab as \( \left( \sigma T^4 / N \right) [1 - \exp(-k_a l)]. If P_1 \) is inside the layer, a scattering direction \( \tilde{u}_2 \) is sampled according to the scattering phase function: \( p_{\Omega_\Sigma}(\tilde{u}_2; \tilde{u}_1). A new scattering path length \( \sigma_2 \) is then sampled and from \( P_1 \) in the direction \( \tilde{u}_2 \) the next scattering location \( P_2 \) can be defined, etc. As before, when the bundle finally leaves the layer (either transmitted or reflected), the fraction of its
initial power that was absorbed within the slab is computed as $(\sigma T^4/N)[1 - \exp(-k_a l)]$ where $l$ is now the total length of the multi-diffusion path.

This algorithm is quite simple in the sense that it is physically intuitible, but the underlying integral formulation is cumbersome as there is no limit to the number of possible scattering events before exit. This means that the vector $x$ of all sampling events required for Monte Carlo weight computation is formally of infinite dimension:

$$x = (\tilde{u}_1, \sigma_1, \tilde{u}_2, \sigma_2, \tilde{u}_3, \sigma_3, \ldots).$$

Practically speaking, however, Eq. (9) may be used directly without entering this formalization exercise. First, the Monte Carlo weight is without ambiguity

$$w(x; k_a) = \sigma T^4[1 - \exp(-k_a l)]$$

(11)

and the probability density function of any bundle path with $n$ scattering events before the layer exit is the product of successive emission and scattering probability density functions:

$$p(x; k_s) = p_{\sigma, e}(\tilde{u}_1)p_{\sigma_2}(\sigma_1; k_s)p_{\sigma, s}(\tilde{u}_2; \tilde{u}_1)p_{\sigma_2}(\sigma_2; k_s)
\cdots
p_{\sigma, s}(\tilde{u}_n; \tilde{u}_{n-1})p_{\sigma_n}(\sigma_n; k_s),$$

(12)

where $p_{\sigma, e}(\tilde{u}_1)$ is the angular probability density function of the emission direction $\tilde{u}_1$, $p_{\sigma_2}(\sigma_j; k_s)$ is the probability density function of the scattering mean free path at the $j$th scattering event, and $p_{\sigma, s}(\tilde{u}_j; \tilde{u}_{j-1})$ is the angular probability density function of the $j$th scattering direction $\tilde{u}_j$ knowing that the incident direction was $\tilde{u}_{j-1}$.

For estimation of the sensitivities to the absorption coefficient, we follow the same conditions as in the preceding example: the sampling probability density functions are independent of $k_a$ and Eq. (9) can be used with

$$\partial_{k_a} w(x; k_a) = l \sigma T^4 \exp(-k_a l)$$

(13)

and

$$\frac{\partial_{k_s} p(x; k_s)}{p(x; k_s)} = 0.$$  

(14)

For estimation of the sensitivities to the scattering coefficient, they are the contrary to the sensitivities of the absorption coefficient: the Monte Carlo weight is insensitive to $k_s$ and therefore only the
Fig. 4. Fraction of the surfacic flux ($A$) emitted by a blackbody that is absorbed by an adjacent, uniform, isotropically scattering slab of thickness $z$, absorption coefficient $k_a$ and scattering coefficient $k_s$. The absorption optical thickness $k_a z$ is fixed to unity and the scattering optical thickness $k_s z$ is varied from 0 (which corresponds to Section 3) to 100. Absorbed fractions (+) correspond to the right axis, whereas the sensitivities to $k_a$ ($\ast$) and $k_s$ ($\times$) correspond to the left axis.

correction term remains:

$$\hat{c}_{k_s} w(x; k_a) = 0$$

and thanks to logarithmic derivations, using the fact that $p_{\Omega_c}$ and $p_{\Omega_s}$ are independent of $k_s$,

$$\frac{\hat{c}_{k_s} p(x; k_s)}{p(x; k_s)} = \hat{c}_{k_s} p_{\Omega_c}(\tilde{u}_1) \frac{p_{\Omega_c}(\tilde{u}_1)}{p_{\Omega_s}(\tilde{u}_1)} + \hat{c}_{k_s} p_{\Omega_s}(\tilde{u}_2; \tilde{u}_1) + \cdots$$

$$= \sum_{j=1}^{n} \left( \frac{1}{k_s} - \sigma_j \right) = \frac{n}{k_s} - \sum_{j=1}^{n} \sigma_j.$$  (16)

Fig. 4$^1$ displays results obtained for both sensitivities with a strict application of Eq. (9) for a case of isotropic scattering using $10^6$ bundle generations.

The same methodology may be used, in the case of anisotropic scattering, to compute sensitivities to the parameters of the scattering phase function. Sensitivities to the asymmetry factor $g$ are displayed in Fig. 5$^2$ using a Henyey–Greenstein phase function [12]. According to this phase function model, cosine $\mu$ of the angle $\chi$ between the incident and scattering directions is distributed according

---

$^1$ For a fixed value of the absorption coefficient, the absorbed flux decreases with increasing scattering coefficient. The reason is that when $k_s$ increases, more photons are reflected after very short distances. $\hat{c}_{k_s} A$ is therefore negative. The fact that the sensitivity of the absorbed flux to $k_s$ is positive is trivial. Its decrease with increasing $k_s$ may be explained by considering again the fact that when $k_s$ increases more photons are reflected after very short distances, reducing the possibility of absorption.

$^2$ A simple physical interpretation of these results may be obtained by considering the medium reduced scattering coefficient $k'_s = (1-g)k_s$ that allows the establishment of an approximate equivalence between an anisotropically scattering medium and an isotropically scattering medium with a lower scattering coefficient: increasing $g$ at constant $k_s$ can be seen as equivalent to decreasing $k'_s$ in an isotropically scattering medium. The physical interpretations proposed for Fig. 4 can therefore be directly translated for interpreting the increase of $A$ with increasing $g$ in the present figure.
Fig. 5. Fraction of the surface flux emitted by a blackbody that is absorbed by an adjacent, uniform, anisotropically scattering slab of thickness \( z \), absorption coefficient \( k_a \) and scattering coefficient \( k_s \). Absorption optical thickness \( k_a z \) and scattering optical thickness \( k_s z \) are fixed to unity. Anisotropic phase functions are represented with Henyey–Greenstein model and the asymmetry factor \( g \) is varied from 0 (isotropic scattering) to 1 (forward scattering). Absorbed fractions (+) correspond to the right axis, whereas the sensitivities to \( g(*) \) correspond to the left axis. Error bars correspond to the estimated standard deviation of the Monte Carlo estimate.

Note 1. On these last simulation results, one may note that the flux and its sensitivity to \( g \) are associated with very distinct statistical relative uncertainties. As a matter of fact, for \( g \) close to 1, Monte Carlo computation of sensitivities to \( g \) is unpractical. Generally speaking, the reason is that the Monte Carlo weights corresponding to \( A \) and \( \partial A / \partial g \) may have very distinct statistical distributions (see appendix). In such extreme cases, it may be impossible to optimize the sampling probability density functions for simultaneous efficient computation of the addressed quantity, on the one hand, and its sensitivities, on the other hand.

Note 2. Another parametric study concerning this physical problem could be to analyze sensitivities to the slab thickness \( z \). On the basis of Eqs. (11) and (12), one could be led to estimate that both the Monte Carlo weight and the sampling probability density function are insensitive to \( z \). The conclusion would be that slab absorption is insensitive to slab thickness, which is an obvious nonsense. The reason is that the space of all possible optical paths changes with slab thickness. Generally speaking,
if the integration domain $D$ depends on the considered parameter $\lambda$, Eqs. (3) and (8) are incorrect because the partial derivation may not be reported inside the integral. In such cases, reformulation exercises may be required and the main interest of the simple above-presented methodology (low-cost implementation of sensitivity procedures) is lost.

6. Conclusion

Our purpose was to show that sensitivity estimations are easy to implement in most Monte Carlo algorithms, provided that they are based on an underlying multiple integral formulation, even if this formulation is not explicit. We used radiative transfer examples for illustration but this approach can be extended to a wider range of applications. If the sampling pdfs are independent of the considered parameter, it is sufficient simply to compute the sensitivity of the Monte Carlo weight and take the average. If the sampling pdfs do depend on the considered parameter, a correction term must be added that is simply the product of the Monte Carlo weight with the logarithmic derivative of the sampling pdf (or the sum of the logarithmic derivative of the pdfs of the successive sampling events).

Appendix. Statistical uncertainties

The numerical behavior of a Monte Carlo algorithm is commonly appreciated referring to the standard deviation of the estimator. With the notations of Section 4, $A = \int_D f(x, \lambda) \, dx = \int_D w(x; \lambda) \, p(x; \lambda) \, dx$ is estimated with any realization of the random variable $S$ defined as the average of $N$ independent realizations of $W$

$$A \simeq s = \frac{1}{N} \sum_{i=1}^{N} w_i.$$  

The standard deviation of $S$ is related to the standard deviation of $W$ according to

$$\sigma_S = \frac{1}{\sqrt{N}} \sigma_W.$$  

Similarly for the sensitivity $\partial_{\lambda} A$, the estimation is made with any realization of the random variable $R$ defined as the average of $N$ independent realizations of $V$

$$\partial_{\lambda} A \simeq r = \frac{1}{N} \sum_{i=1}^{N} v_i$$

with

$$v_i = \partial_{\lambda} w(x_i; \lambda) + w(x_i; \lambda) \frac{\partial_{\lambda} p(x_i; \lambda)}{p(x_i; \lambda)}.$$  

The standard deviation of $R$ is related to the standard deviation of $V$ according to

$$\sigma_R = \frac{1}{\sqrt{N}} \sigma_V.$$
Both $\sigma_S$ as $\sigma_R$ tend to zero when increasing $N$ to infinity, which ensures the convergence of the Monte Carlo algorithm, but the convergence speed is entirely driven by the values of $\sigma_W$ and $\sigma_V$, respectively. The choice of the sampling probability density function $p$ is therefore essential to ensure fast convergence (small number of required sampling events) as it fixes the distributions of the Monte Carlo weights $W$ and $V$.

One of the advantages of the presented methodology is that the same sampling events may be used both for estimation of $A$ and $\partial_A A$. The same sampling probability density function $p$ must therefore be used for $W$ and $V$. Consequently, the question arises of the possibility that $p$ be adjusted in such a way that both $\sigma_W$ and $\sigma_V$ are satisfactory.

There is no systematic answer to this question. Generally speaking, detailed understanding of the physical processes at work allows successful optimization of the integral evaluation [13]. This does not necessarily ensure that the sensitivity estimation is optimized. For problems in which the sensitivity estimation is essential, similar physical reasoning may be concentrated on the sensitivity itself. But as illustrated in the following example, there is no mathematical constraint relating the statistics of $W$ and $V$.

Consider the case where the function to be integrated is

$$f(x; \lambda) = \alpha + \lambda g(x),$$

where $\alpha$ is a constant and $g$ is any function of $x$ independent of the parameter $\lambda$. In the particular case $\lambda = 0$, $f$ becomes independent of $x$. If $p$ is taken as uniform on $\mathcal{D}$, $w(x; 0) = f(x; 0)/p$ is independent of $x$ and therefore $\sigma_W = 0$. This corresponds to an ideal optimization of $A$ estimate. Correspondingly, $v(x; 0) = g(x)/p$ and the standard deviation $\sigma_V$ can take any value depending on $g$.

References