Laboratori Nazionali di Frascati

LNF-69/49

R. Baldini-Celio, B. Ballico-Lay and M. A. Spano-Mencuccini :
A MULTISTAGE SELF-IMPROVING MONTE CARLO METHOD

Estratto da : Nuclear Instr. and Meth. 72, 317 (1969)

NUCLEAR INSTRUMENTS AND METHODS 72 (1969) 317–320; © NORTH-HOLLAND PUBLISHING CO

A MULTISTAGE SELF-IMPROVING MONTE CARLO METHOD

R. BALDINI-CELIO, B. BALLICO-LAY and M. A. SPANO-MENCUCCINI

Laboratori Nazionali di Frascati del CNEN, Frascati, Italy

Received 14 January 1969

We describe a technique to speed up the application of the Monte Carlo method to calculate the detection efficiency of a physical process by an experimental apparatus. The calculation goes through more stages. At each stage the Monte Carlo obtains informations, by its own results, which allows itself to reduce the variance in the next stage by means a weighted sampling technique. Then it is possible to achieve the same results as a crude Monte Carlo in a running time about 100 times shorter.

1. Introduction

In this report we describe a technique (self-improving Monte Carlo) to speed up the application of the Monte Carlo method (M.C.) to calculate the detection efficiency of a physical process by an experimental apparatus*.

As is known, the advantages of the M.C. method are many and remarkable. It is possible, for instance, to stop and to resume the calculation at any time, and to achieve a simultaneous estimate of the efficiency and its standard error.

On the other side, the very slow convergence may be considered a great obstacle in the application of the above method. Often one needs several running hours of a big computer if the efficiency involved is small and no particular technique is used to simplify the calculation.

Usually these techniques consist in bounding the range of variation of the quantities randomly selected by the M.C., in order not to simulate events a priori undetectable, and then correcting the efficiency value by a suitable factor.

Such a procedure requires some knowledge of the process in order to fix such limitations, e.g. kinematic cuts. But frequently this knowledge is not readily available.

In the technique we are going to describe this knowledge comes out from the results obtained, by the M.C. itself, in a previous stage of the calculation. This knowledge is used, by the M.C. itself, to concentrate trials in the regions regarded as the most suitable. As it will be seen below, by a multistage technique, it is possible to achieve the same results as a crude M.C. in a running time about 100 times shorter.

* This calculation usually is made by simulating many times at random the physical process of interest. At the end of each simulated process, we test if the event has the requisites to be detected by our experimental apparatus. The efficiency is defined as: $\varepsilon = N_d/N_s$, where N_d is the number of detected events and N_s is the number of simulated events. Although the technique we describe is quite general, we have applied it here to the detection efficiency of a reaction among elementary particles.

2. Weighted sampling technique

In this section, a description of "weighted sampling" technique is given. This technique is a variant of the standard methods[1]) "importance sampling" and "stratified sampling".

Let us define V_0 as the volume of the random quantities picked out by the M.C.

Let us divide V_0 into m volumes V_i and attribute to each V_i a weight p_i (which we shall fix later).

Then we pick out a volume V_i with its probability p_i and, inside V_i, we pick out an event according to a crude M.C.

Let us call n_i the number of detected events belonging to V_i and N the total number of simulated events.

Then the efficiency ε and its standard error σ^2 are:

$$\varepsilon = \lim_{N\to\infty} \sum_{i=1}^{m} \{V_i/(p_i V_0)\}(n_i/N), \tag{1}$$

$$\sigma^2 = \lim_{N\to\infty} (1/N) \left\{ \sum_{i=1}^{m} \{V_i/(p_i V_0)\}(n_i/N) - \left[\sum_{i=1}^{m} \{V_i/(p_i V_0)\}(n_i/N) \right]^2 \right\}. \tag{2}$$

In fact, the distribution of the events in V_i being uniform, the efficient part of V_i is $\varepsilon_i V_0$. Therefore, the probability for an event, belonging to V_i and being detectable is:

$$q_i = p_i(\varepsilon_i V_0/V_i).$$

Hence:

$$\varepsilon = \sum_{i=1}^{m} \{V_i/(p_i V_0)\} q_i = \lim_{N\to\infty} \sum_{i=1}^{m} \{V_i/p_i V_0\}(n_i/N).$$

Eq. (2) may be easily obtained by observing that q_i is a polynomial distribution.

The interesting point now is that we may reduce the variance by a suitable choice of p_i and V_i. For instance, if we choose:

$$p_i = \varepsilon_i/\varepsilon,$$
$$V_i = \varepsilon_i V_0, \qquad (3)$$

we have:

$$\sigma^2 = 0.$$

The conditions (3) are obviously not realizable, because we do not know the efficient volumes previously. However, the variance will be greatly reduced by choosing (as we shall do) weights and volumes approaching to eq. (3).

3. Multi-stage self-improving M.C.

Generally the multi-stage self-improving M.C. operates as follows: at the end of each stage, the events detected and already in memory are grouped into volumes V_i. A weight p_i is consequently given, according to the number of events detected belonging to V_i. In the further stage volumes V_i are sampled according to their weights. Weights and volumes so obtained will represent a sequence which tends asymptotically to eq. (3).

In any case this technique leads to a reduction of the running time. However, this method, as it has been described, is not very profitable when the efficiency we want to estimate is very small. In fact, in this case, the information necessary to construct weights and volumes would require, at least for the first stage, a very long running time.

Therefore it is advisable to increase the information that the M.C. can obtain from the events already selected.

In fact we remark that the efficiency is strongly dependent upon the parameters which define the resolution power of the experimental apparatus (such as, for instance, the solid angle subtended by a detector).

Generally a small increase in the range of variation of these parameters (control parameters) causes a remarkable increase in the detection efficiency without strongly altering the kinematic situation.

This fact suggests a procedure for speeding up the self-improving M.C. We increase the range of variation of some of the control parameters. Then we record the events detected under these conditions and, making use of these events, we construct weights and volumes for the next stage.

At every stage we decrease the range of variation of the control parameters, from an initial value to the correct one, and, as the stages go on, weights and volumes will converge rapidly to the conditions (3).

* Obviously one must know, at least approximately, the efficiency as a function of the range of variation of the control parameters.

On this basis, we found it advisable to choose the following weights and volumes at the $(j+1)^{\text{th}}$ stage ($\varepsilon_i^{(j)} V_0$ is the efficient volume V_i as estimate at the j^{th} stage):

$$V_i = \varepsilon_i^{(j)} V_0,$$
$$p_i = \varepsilon_i^{(j)} \{p_0 / \sum_k \varepsilon_k^{(j)}\}. \qquad (3a)$$

Here $1-p_0$ is a weight which can be ascribed to the volume $V_0 - \sum_k V_k$, $\sum_k V_k$ being the sum of the volumes for which $\varepsilon_i^{(j)} \neq 0$. It is a free parameter which can be fixed on the basis of the required final accuracy [formula (4)].

With choices (3a) the distribution of events in the volume assumed efficient is also assumed uniform. This fact may be useful because we do not know, where the events detected in the following stage will go, within this volume.

However, it is obvious that a wrong choice of weights does not change the efficiency value. The only consequence of a wrong choice of weights might be a reduction in optimization.

4. Variance and "optimization factor" for a multi-stage M.C.

Now, we introduce the quantity k_j defined as the ratio:

$$k_j = \varepsilon^{(j)}/\varepsilon,$$

where $\varepsilon^{(j)}$ is the efficiency of the j^{th} stage*, and ε is the efficiency we want to estimate.

To make things easier, we suppose there are no events in the region $V_0 - \sum_k V_k$, defined above.

If we choose the weights according to the last section rules, the j^{th} stage variance is (N_j is the number of simulation of the j^{th} stage):

$$\sigma_j^2 \simeq \{(k_j\varepsilon)^2/N_j\}[\{k_{j-1}/(k_j p_0)\} - 1], \qquad j = 2,\ldots,f.$$

If the efficiency varies rapidly from stage to stage, the following inequality holds:

$$k_{j-1}/k_j \gg 1,$$

hence:

$$\sigma_j^2 \simeq \varepsilon^2\{k_j k_{j-1}/(N_j p_0)\}, \qquad j = 2,\ldots,f.$$

The first stage is a crude M.C., therefore:

$$\sigma_1^2 \simeq (k_1\varepsilon/N_1)(1-k_1\varepsilon) \simeq k_1\varepsilon/N_1.$$

Let us fix the accuracy with which we want to estimate the efficiency at the j^{th} stage; i.e. we require:

$$\sigma_j/(k_j\varepsilon) \simeq c, \qquad j = 1,\ldots,f-1.$$

Then the final state variance will be:

$$\sigma_f^2 \simeq k_{f-1}\varepsilon^2/(N_f p_0) \simeq \varepsilon/\{(p_0c^2)^{f-1}N_1 \ldots N_f\}.$$

In order to compare the self-improving M.C. and the crude M.C., we introduce the "optimization factor" F_{opt}. F_{opt} is the ratio between the running time necessary to achieve the same variance (or also, the running time being the same, the ratio between the variances).

If N is the total number of simulations, we have for crude M.C.

$$\sigma^2_{crude} = \varepsilon(1-\varepsilon)/N \simeq \varepsilon/N.$$

As one regards the self-improving M.C. the increase of running time due to the "weighted sampling" is negligible. Therefore, the comparison may be made for an equal number of simulations.

We get, finally:

$$F_{opt} = \sigma^2_{crude}/\sigma^2_{opt} \simeq \{(p_0c^2)^{f-1}N_1 \ldots N_f\}/N. \quad (4)$$

Bearing in mind that

$$\sum_{i=1}^{f} N_i = N,$$

the condition which leads to a maximum of the optimization is:

$$N_1 = N_2 = \ldots = N_f = N/f, \quad (5)$$

when we fix the total number of simulations, the accuracy required to proceed to a new stage, the weight p_0 and the number of stages.

The "optimization factor" as a function of the number of the stages has a maximum for:

$$n_0 = (p_0c^2N)/\mathrm{e}.$$

Therefore

$$F_{opt\ max} = \mathrm{e}^{n_0}/(p_0c^2N). \quad (6)$$

5. Limitations and practical applications of the multi-stage self-improving M.C.

If we substitute reasonable values of c, N and p_0 in eq. (6), we obtain exceedingly high optimization factors. However, they are not realistic because conditions leading to the maximum of the optimization are not realizable in practice.

For instance, one cannot choose arbitrarily the range of variation of the control parameters at the first stage. In fact if we increase them too much, the M.C. may pick kinematic conditions too different from the required ones, thus selecting regions which in the end will be not efficient at all.

The range of variation of the control parameters referring to consecutive stages must be so different that the region actually efficient, at the $(j+1)^{th}$ stage, is wholly contained in the interior of the region assumed as efficient on the basis of the j^{th} stage*.

All the previous conditions practically limit the number of stages which may be derived from the equations of section 4.

A general rule for the best choice of the parameters of the self-improving M.C. is difficult to give. A possible procedure is the following: we fix the range of variation of the control parameters at the first stage and the percentage statistical error at every stage preceding the last one. From the running time at our disposal we can derive the total number of events to be simulated. We fix then the number of stages which allows the differences between control parameters referring to consecutive stage to be great enough.

To give a crude example, we quote the optimization factor expected for the calculation of an efficiency $\varepsilon \approx 10^{-4}$ carried out by a two-stage self-improving M.C. with an equal number of simulations of each stage, according to the rule (5).

If we assume $c = 5\%$, $N = 10^5$ hence $k_1 \approx 100$, we get:

$$F_{opt} \approx 100.$$

As we may see, even though this value is far from the maximum factor we may theoretically obtain, it is already sufficient for the practical purposes of a rough improvement on a M.C.

6. An application of the multi-stage self-improving M.C.

As a check we calculated by the self-improving M.C. the detection efficiency of the reaction:

$$\gamma + \mathrm{p} \rightarrow \eta + \mathrm{p}$$
$$\qquad\qquad \downarrow \rightarrow 2\gamma,$$

observed by two total absorption Čerenkov detectors which detect the two γ decay of the η-meson[2]).

The solid angle $\Delta\Omega_c$ subtended by each Čerenkov detector has been chosen as control parameter to be varied at each stage. In fact, we have in first approximation:

$$\varepsilon \propto (\Delta\Omega_c)^2.$$

The calculation has been carried out by a two- or three-stage self-improving M.C. At the first stage each $\Delta\Omega_c$ has been chosen to be 16 times greater than the solid angles used in the experiment ($\simeq 4°$). The percentage

* If this condition is achieved one can put $p_0 = 1$. Moreover it allows us to define volumes V_i for the next stage, beginning from the recorded events n_i, with some degree of freedom.

TABLE 1

Number of stages	Total number of simulated events N	$(\varepsilon \pm \sigma_{opt}) \cdot 10^4$	$F_{opt} \simeq \varepsilon / \{N(\sigma_{opt})^2\}$
2	6×10^4	0.29 ± 0.03	50
	9×10^4	0.28 ± 0.03	35
3	9×10^4	0.32 ± 0.02	80
	36×10^4	0.28 ± 0.01	80

accuracy required to proceed to a new stage was $\approx 5\%$ and we assumed $p_0 = 1$.

The calculation of this detection efficiency has already been carried out by a crude M.C. where kinematic cuts were introduced. The efficiency values calculated by the self-improving M.C. may therefore be checked. For the sake of generality in the self-improving M.C., kinematic cuts were not introduced; therefore the optimization factor quoted refers to a crude M.C. without kinematic cuts (however, if these cuts are introduced in the self-improving M.C., the optimization factor does not change).

We emphasize that the self-improving M.C. technique is advantageous mainly when these cuts may not be fixed previously.

Results obtained by the self-improving M.C. are quoted in table 1. The efficiency check, calculated as above, is

$\varepsilon = (0.30 \pm 0.01) \cdot 10^{-4}$ {detected/simulated} events.

The authors are particularly grateful to Prof. A. Reale for his helpful criticism and supervision in the present work.

References

[1] J. M. Hammersley and D. C. Handscomb, *Monte Carlo methods* (Methuen's monographs);
Y. Shreider, *Method of statistical testing* (Elsevier Publ. Company).

[2] C. Bacci, R. Baldini-Celio, C. Mencuccini, A. Reale, M. Spinetti and A. Zallo, Phys. Rev. Letters **20** (1968) 11.