Adaptive Monte Carlo for nuclear data evaluation

Georg Schnabel[a]

Irfu, CEA, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

Abstract. An adaptive Monte Carlo method for nuclear data evaluation is presented. A fast evaluation method based on the linearization of the nuclear model guides the adaptation of the sampling distribution towards the posterior distribution. The method is suited for parallel computation and provides detailed uncertainty information about nuclear model parameters. Especially, the posterior distribution of the model parameters is not restricted to be multivariate normal. The method is demonstrated in an evaluation of the ^{181}Ta total cross section for incident neutrons. Future applications are as an efficient sampling scheme in the Total Monte Carlo method, and the restriction of parameter uncertainties in nuclear models by both differential and integral data.

1. Introduction

Computer simulations guide the development of nuclear facilities, such as fusion reactors and accelerator-driven systems for nuclear waste incineration. They provide estimates of observables, such as production rates of atomic elements and neutron multiplication factors, which help optimizing design characteristics with respect to efficiency and safety. An important prerequisite to perform a simulation is evaluated nuclear data in the form of an ENDF file. If this ENDF file contains besides cross sections and related quantities also the associated uncertainties as covariance matrices, the latter can be propagated through the simulation by means of perturbation theory. The additional information about uncertainties in the simulation results is of great value. For example, one wants to know whether an obtained neutron multiplication factor of 0.95 could also exceed one due to uncertainties of data in the ENDF file.

The propagation of covariance matrices using perturbation theory, however, exhibits a fundamental shortcoming. Both the uncertainties of evaluated nuclear data and the uncertainties of simulation results are restricted to be of Gaussian shape. The non-linearity of nuclear models and non-linear features in simulations may lead to distributions that have several peaks and are very skewed. Replacing such distributions by Gaussian distributions may mean to underestimate the probability of rare events or to wrongly regard domains in parameter space lying in-between two peaks of probability as highly likely.

In former times, lacking any alternative, this shortcoming has been ignored. More recently, advances in computing technology facilitate the application of Monte Carlo methods. Monte Carlo methods (e.g. [1,2]) do not replace nuclear models by linearized versions and are not dependent on the assumption of Gaussian distributions. Mean values, uncertainties (i.e., standard deviations), and other important quantities are recovered from a sample of model parameter sets and corresponding model predictions. Within the framework of the Total

Monte Carlo method [3], even the simulation of a nuclear facility is performed for each parameter set. Just as for a sample of nuclear model predictions, mean values and standard deviations can also be calculated from a sample of simulation results.

Despite the practical success of Monte Carlo methods, they are very resource-intensive and orders of magnitude slower than evaluation methods relying on the linearization of nuclear models. It is therefore of utmost importance to use an efficient sampling scheme in order to accelerate convergence and, related, to reduce execution time. For this reason, we introduce an adaptive Monte Carlo scheme that incorporates a fast evaluation scheme based on the linearization of nuclear models to adapt the sampling distribution towards the posterior distribution. This measure can significantly speed up convergence, as will be demonstrated in an evaluation of the ^{181}Ta total cross section for incident neutrons.

2. Method

We briefly introduce the formulas used in nuclear data evaluation together with the formulas for Monte Carlo integration. The adaptive Monte Carlo method is presented thereafter in Sect. 2.2.

2.1. Preliminaries

The objective of Bayesian evaluation methods is to extract information from the so-called *posterior distribution*

$$\pi(\vec{p} \,|\, \vec{\sigma}_{\exp}) \propto \mathcal{N}(\vec{\sigma}_{\exp} \,|\, \mathcal{M}(\vec{p}), \mathbf{B}) \times \mathcal{N}(\vec{p} \,|\, \vec{p}_0, \mathbf{A}_0), \quad (1)$$

which combines information from experiments, given as a measurement vector $\vec{\sigma}_{\exp}$ and associated covariance matrix $\mathbf{B}$, with information from a nuclear model, given as a best parameter vector $\vec{p}_0$ and associated covariance matrix $\mathbf{A}_0$. The notation $\mathcal{N}(\vec{x} \,|\, \vec{x}_0, \mathbf{\Sigma})$ indicates the probability density at location $\vec{x}$ of a multivariate normal distribution centered at $\vec{x}_0$ with covariance matrix $\mathbf{\Sigma}$. The functional form of a multivariate normal distribution is

---

[a] e-mail: `georg.schnabel@cea.fr`

given by

$$\mathcal{N}(\vec{x} \mid \vec{x}_0, \mathbf{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d \det \mathbf{\Sigma}}}$$
$$\times \exp\left\{-\frac{1}{2}(\vec{x} - \vec{x}_0)^T \mathbf{\Sigma}^{-1}(\vec{x} - \vec{x}_0)\right\} \quad (2)$$

with d being the dimension of $\vec{x}$. In Eq. (1), $\mathcal{M}(\vec{p})$ denotes the prediction of the nuclear model based on parameter set $\vec{p}$ for the observables in $\vec{\sigma}_{\text{exp}}$. In other equations, $\mathcal{M}(\vec{p})$ shall denote the prediction of the observable that fits into the context. Due to the complexity of nuclear models, the normalization constant in Eq. (1) is usually not available in closed-form and has to be estimated by Monte Carlo integration.

Important statistics, such as the expectation of a cross section $\sigma = \mathcal{M}(\vec{p})$ and the associated variance (=standard deviation squared), can be extracted from the posterior distribution $\pi(\vec{p} \mid \vec{\sigma}_{\text{exp}})$,

$$\mathbb{E}\left[\mathcal{M}(\vec{p})\right] = \int \mathcal{M}(\vec{p})\, \pi(\vec{p} \mid \vec{\sigma}_{\text{exp}})\, d\vec{p}, \quad (3)$$

$$\text{Var}\left[\mathcal{M}(\vec{p})\right] = \int (\mathcal{M}(\vec{p}) - \mathbb{E}[\sigma])^2\, \pi(\vec{p} \mid \vec{\sigma}_{\text{exp}})\, d\vec{p}. \quad (4)$$

Unfortunately, these integrals are usually intractable because of the nuclear model appearing therein. The long execution time of model codes and the large dimension of the parameter vector prevent the application of direct numerical integration methods. Therefore, Monte Carlo integration is used instead, which will be outlined in the following paragraphs.

In Monte Carlo integration, we have to specify a *sampling distribution* $\tau(\vec{p})$, which should closely resemble the posterior distribution $\pi(\vec{p} \mid \vec{\sigma}_{\text{exp}})$, and from which we need to be able to obtain samples. Ideally, the sampling distribution is identical to the posterior distribution for maximal efficiency. In practice, this desirable agreement cannot be reached due to the complexity of nuclear models.

A certain number N of parameter vectors $\vec{p}_i$ is then drawn from the sampling distribution. The model prediction is evaluated for each $\vec{p}_i$ which enables the calculation of the weights

$$\omega_i = \frac{\pi(\vec{p}_i \mid \vec{\sigma}_{\text{exp}})}{\tau(\vec{p}_i)}. \quad (5)$$

Finally, the values of integrals, such as in Eqs. (3) and (4), can be estimated by

$$\mathbb{E}\left[f(\vec{p})\right] \approx \frac{\sum_{i=1}^{N} \omega_i f(\vec{p}_i)}{\sum_{i=1}^{N} \omega_i}. \quad (6)$$

This formula works for the important case where the posterior distribution is not normalized. In this case, the expected value of the denominator is N/C with C being the normalization constant of the posterior distribution.

If the sampling distribution is very different from the posterior distribution, the values of the weights are spread over many orders of magnitude. Consequently, only a few terms in Eq. (6) contribute to the estimate, thereby decreasing its accuracy. The notion of decrease in accuracy

can be quantified by a measure called the *effective sample size* [4],

$$N_{\text{eff}} = \frac{(\sum_{i=1}^{N} \omega_i)^2}{\sum_{i=1}^{N} \omega_i^2}. \quad (7)$$

If we are in the lucky position to be able to directly draw samples from the posterior distribution, the effective sample size equals the number of drawn samples N. In the other extreme case of very bad alignment between sampling and posterior distribution, the effective sample size drops to one.

A common choice for the sampling distribution is the prior distribution $\mathcal{N}(\vec{p} \mid \vec{p}_0, \mathbf{A}_0)$ appearing in Eq. (1). The weights are then determined by the likelihood $\mathcal{N}(\vec{\sigma}_{\text{exp}} \mid \mathcal{M}(\vec{p}), \mathbf{B})$. If the posterior distribution is mainly characterized by the likelihood due to a lot of precise measurements, the posterior distribution is much more peaked than the prior distribution. Then, despite a large number of samples, we gain essentially no information about the posterior distribution due to a very low effective sample size.

2.2. Adaptive Monte Carlo method

To obtain a sampling distribution well adapted to the posterior distribution, we propose an adaptive Monte Carlo scheme that refines the sampling distribution in the course of the sampling process. The sampling distribution is chosen to be a mixture of multivariate normal distributions,

$$\tau(\vec{p}) = \sum_{k=1}^{K} \alpha_k \mathcal{N}(\vec{p} \mid \vec{x}_k, \mathbf{\Sigma}_k), \quad (8)$$

from which samples can be efficiently generated.

The sampling process is organized in stages performed one after another. Let s be the index differentiating between stages. Because the sampling distribution may be updated after a stage, let $\tau_s(\vec{p})$ denote the sampling distribution that has been used to generate the parameter vectors of stage s. Using this notation, we now describe the steps of the adaptive Monte Carlo method:

1. Initialize the stage counter, $s := 1$, and set the sampling distribution to the prior distribution, $\tau_1(\vec{p}) := \mathcal{N}(\vec{p} \mid \vec{p}_0, \mathbf{A}_0)$.
2. Generate a number N of parameter vectors $\{\vec{p}_{si}\}_{i=1..N}$ from the current sampling distribution $\tau_s(\vec{p})$, compute the associated model predictions $\mathcal{M}(\vec{p}_{si})$, and on their basis the weights ω_{si} according to Eq. (5).
3. Use the weights ω_{si} of the current stage to calculate the effective sample size according to Eq. (7). If the effective sample size is below a user-defined threshold, do a learning step to update the sampling distribution, i.e., $\tau_{s+1} \neq \tau_s$. Otherwise, set $\tau_{s+1} := \tau_s$. The learning step will be discussed separately after this enumeration.
4. Use the weights of all stages so far to calculate the total effective sample size according to Eq. (7). If it is smaller than a user-specified target value, set $s := s + 1$ and continue with step 2. Otherwise continue with step 5.
5. Compute estimates for observables of interest by taking recourse to Eq. (6) and using the weights, parameter vectors, and model predictions of all stages.

2

The following description of the learning step completes the algorithm. The K parameter vectors $\vec{p}_k$ with the highest weights are selected from the current stage. The proper choice of the number K will become apparent in a moment. For each of these parameter vectors, we construct a linear model

$$\mathcal{M}_{\text{Lin}}^k(\vec{p}) = \mathcal{M}(\vec{p}_k) + \mathbf{S}_k(\vec{p} - \vec{p}_k). \qquad (9)$$

The matrix $\mathbf{S}_k$ is the Jacobian matrix containing the partial derivatives of the model predictions with respect to the model parameters. In other words, we generate linear Taylor approximations of the nuclear model by expanding around the vectors $\vec{p}_k$.

Assuming that one model calculation provides a prediction for all observables of interest, the determination of the Jacobian matrix requires $\#P + 1$ model calculations where $\#P$ means the number of model parameters. Therefore, if we are able to perform N calculations in parallel, we can construct $K = N/(\#P + 1)$ linear models at the same time.

The posterior distribution in Eq. (1) is proportional to the product of two multivariate normal distributions. If we use a linear model $\mathcal{M}_{\text{Lin}}^k$, also the associated posterior distribution is multivariate normal. Its center vector $\vec{p}_{\text{ev},k}$ and covariance matrix $\mathbf{A}_{\text{ev},k}$,

$$\vec{p}_{\text{ev},k} = \vec{p}_0 + \mathbf{A}_0 \mathbf{S}_k^T (\mathbf{S}_k \mathbf{A}_0 \mathbf{S}_k^T + \mathbf{B})^{-1} (\vec{\sigma}_{\text{exp}} - \mathcal{M}_{\text{Lin}}^k(\vec{p}_0)), \qquad (10)$$

$$\mathbf{A}_{\text{ev},k} = \mathbf{A}_0 - \mathbf{A}_0 \mathbf{S}_k^T (\mathbf{S}_k \mathbf{A}_0 \mathbf{S}_k^T + \mathbf{B})^{-1} \mathbf{S}_k \mathbf{A}_0, \qquad (11)$$

completely characterize the respective posterior distribution. We evaluate Eqs. (10) and (11) using the linear models of Eq. (9) and obtain K multivariate normal distributions $\mathcal{N}(\vec{p} \mid \vec{p}_{\text{ev},k}, \mathbf{A}_{\text{ev},k})$. This approximate solutions are used to update the sampling distribution of the form specified in Eq. (8),

$$\tau_{s+1}(\vec{p}) = \tau_s(\vec{p}) + \sum_{k=1}^{K} \beta_k \mathcal{N}(\vec{p} \mid \vec{p}_{\text{ev},k}, \mathbf{A}_{\text{ev},k})$$

$$= \sum_{j=1}^{J} \alpha_j \mathcal{N}(\vec{p} \mid \vec{x}_j, \mathbf{\Sigma}_j)$$

$$+ \sum_{k=1}^{K} \beta_k \mathcal{N}(\vec{p} \mid \vec{p}_{\text{ev},k}, \mathbf{A}_{\text{ev},k}). \qquad (12)$$

The new proportions β_k are set to one. Then old and new proportions are rescaled so that they sum up to one. This first assignment is the basis for a more fine-grained adjustment of the proportions to be discussed next.

Each component covers a certain volume in parameter space. For a parameter vector $\vec{p}$, we can calculate the probability that it belongs to a specific component j,

$$m_j(\vec{p}) = \frac{\alpha_j \mathcal{N}(\vec{p} \mid \vec{x}_j, \mathbf{\Sigma}_j)}{\tau_{s+1}(\vec{p})}. \qquad (13)$$

The probability of a parameter vector to belong to a new component β_k is calculated analogously. Using this membership probability, we can express the amount of probability mass of the posterior distribution associated with the jth component by

$$P_j = \int m_j(\vec{p}) \, \pi(\vec{p} \mid \vec{\sigma}_{\text{exp}}) \, \mathrm{d}\vec{p}. \qquad (14)$$

This integral is in general not tractable and has to be approximated by means of Monte Carlo integration,

$$P_j \approx \frac{\sum_{i=1}^{N} \omega_{si} \, m_j(\vec{p}_{si})}{\sum_{i=1}^{N} \omega_{si}}, \qquad (15)$$

with the weights as defined in Eq. (5). Importantly, an unnormalized version $\tilde{\pi}(\vec{p} \mid \vec{\sigma}_{\text{exp}})$ of the posterior distribution can be used.

The weights and parameter vectors are those obtained in the current stage s. Due to the approximation of the integral, which may be very inaccurate if the effective sample size is small, we have to introduce a special rule: A vector $\vec{p}_k$ selected from the vectors $\{\vec{p}_{si}\}_{i=1..N}$ as expansion point for the linear model $\mathcal{M}_{\text{Lin}}^k$ is assigned with probability one to the new component $\mathcal{N}(\vec{p} \mid \vec{p}_{\text{ev},k}, \mathbf{A}_{\text{ev},k})$, or shortly, $m_k(\vec{p}_k) = 1$ and $m_q(\vec{p}_k) = 0, \forall q \neq k$. Some experimentation has shown that this rule is needed for convergence.

Under the assumption that the components are locally well aligned to the posterior distribution, a good assignment for the proportions is $\alpha_j = P_j$. This assignment scales the multivariate normal distributions to locally match the posterior distribution as close as possible. The expression for the β_k is analogous.

Having outlined how new multivariate normal components are introduced in the sampling distribution, and how their proportions and distribution parameters are determined, the learning step is completely defined.

3. Application

We demonstrate the adaptive Monte Carlo method in an evaluation of the ^{181}Ta total cross section for incident neutrons with energies between 5 and 100 MeV. We performed the model calculations with the nuclear model code TALYS [5] which implements the optical model in the global parametrization of Koning and Delaroche [6]. The evaluation was restricted to three of the most sensitive model parameters, namely the real radius r_v, the real diffuseness a_v, and the leading coefficient v_1 in the formula for the real potential depth. A parameter with a tilde on top, such as $\tilde{r}_v$, equals the respective parameter without a tilde divided by the TALYS default value.

We used a normal distribution as prior for these variables with a standard deviation of 10% relative to the TALYS default values. All other parameters were fixed at their default values. The likelihood was determined by the measurements of Finlay and colleagues [7]. Both statistical uncertainties and a normalization error were considered for the data points.

In this schematic evaluation, we were able to obtain an estimate of the posterior distribution, Eq. (1), by means of the Metropolis-Hasting algorithm, see Fig. 1. This approximation serves as benchmark for the adaptive Monte Carlo method.

The adaptive Monte Carlo method has been set up to generate hundred parameter vectors in each iteration. Three linear approximations were constructed at once in each learning step. A learning step was applied when the effective sample size within an iteration dropped below eighty.

The sampling distribution converged rapidly towards the posterior distribution. Even though the effective sample
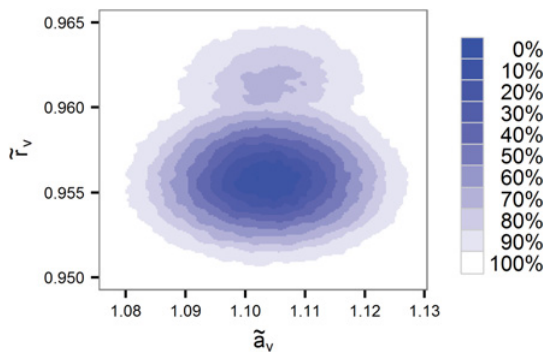
Figure 1. Approximation of the posterior distribution. The different shades of blue indicate the confidence areas.
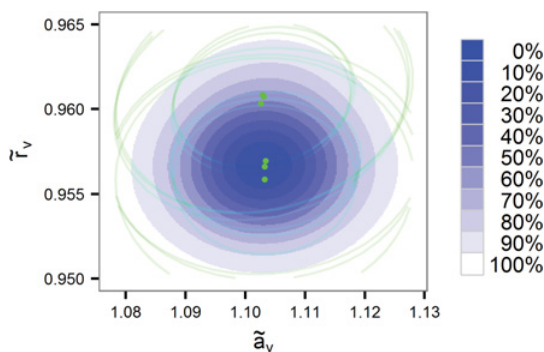
Figure 2. Sampling distribution after ten iterations. The green ellipses indicate the components of the sampling distribution with a proportion greater than one percent.

size was about one in the first iteration, after ten iterations it was usually as high as ninety. Sometimes this efficiency has already been reached after just a few iterations. However, we also remark that an effective sample size larger than 95 has never been observed, even after dozens of iterations corresponding to thousands of acquired samples. This fact shows that the algorithm still has limits in its capability to adapt to the posterior distribution. The impact of this limitation has yet to be studied in evaluations with more adjustable model parameters and different nuclear models.

A representative sampling distribution after ten iterations is depicted in Fig. 2. The comparison with Fig. 1 shows good global agreement between the posterior and the sampling distribution. Noteworthy, also the small local maximum in the top part of Fig. 1 has been captured by some components of the sampling distribution. Thus, the sampling distribution is capable of adapting to a multipeaked posterior distribution.

4. Conclusion

An adaptive Monte Carlo method has been presented. It exploits an evaluation method based on the linearization of the nuclear model to adapt the sampling distribution towards the posterior distribution, which may exhibit several local maxima.

The adaptive Monte Carlo method performed very well in a schematic evaluation of the ^{181}Ta total cross section for incident neutrons with three model parameters. We anticipate the method to work also well with a larger number of parameters if at least one of the following conditions is fulfilled: (1) the likelihood is sharply peaked and constrains the model parameters to linear domains of the model, or (2) the variations of the prediction vector due to variations of the parameters are locally embedded in a low-dimensional vector space. In the absence of these conditions, many learning steps will occur without significantly improving the effective sample size. Future investigations will show for which nuclear models these conditions are satisfied.

Applications within the domain of nuclear data evaluation may be as a part in the Total Monte Carlo method. An extended version may be useful to treat model defects by means of Monte Carlo. Yet another use case could be constraining model parameters by both differential and integral observables. Especially this latter possibility may offer a huge benefit to constrain model parameters of high energy models such as INCL. An extended version of the method is under preparation for publication.

References

[1] A.J. Koning, The European Physical Journal A **51**(12), 184 (2015)
[2] R. Capote, D.L. Smith, A. Trkov, M. Meghzifene, Journal of ASTM International **9**(3), 179 (2012)
[3] A.J. Koning, D. Rochman, Annals of Nuclear Energy **35**(11), 2024 (2008)
[4] A. Kong, Tech. Rep. 348, Department of Statistics, University of Chicago, Chicago (1992)
[5] A.J. Koning, S. Hilaire, S. Goriely, *TALYS-1.6 - A Nuclear Reaction Program*, http://www.talys.eu (2013)
[6] A.J. Koning, J.P. Delaroche, Nuclear Physics A **713**(3–4), 231 (2003)
[7] R. Finlay, W. Abfalterer, G. Fink, E. Montei, T. Adami, P. Lisowski, G. Morgan, R. Haight, Physical Review C **47**(1), 237 (1993)