

Prospects of Matrix Elements Integration with Markov Chain Monte Carlo

Lev Dudko¹ Andrey Popov²

¹SINP MSU ²aa.popov@www-hep.sinp.msu.ru, SINP MSU

QFTHEP 2010, Golitsyno, Moscow, Russia, 14.09.2010

Outline

- ▶ The traditional approach to integrate the squared matrix element and some of its difficulties.
- ▶ What is Markov chain Monte Carlo? Its pros and cons.
- ▶ An illustration on integration with Markov chain Monte Carlo.
- ▶ Summary and conclusions.

This report is rather a call to pay attention to a perspective group of methods than a completed research.

The Traditional Way of Cross-Section Calculation

- ▶ The squared matrix element (SQME) is usually integrated by the means of VEGAS algorithm or other adaptive Monte Carlo (MC) techniques.
- ▶ VEGAS samples the probe points from a piecewise constant coordinate-wise factorable distribution which approximates the original SQME.
- ▶ The better this approximation is, the smaller the uncertainty. But this condition cannot be satisfied easily since the SQME is usually poorly coordinate-wise factorable because of the peak “ridges” corresponding to the propagators’ denominators.
- ▶ To overcome the obstacle one can introduce a smart remapping of the phase space that smooths the “ridges” (“regularization”).

Problems of the Traditional Way

In CompHEP package the phase space remapping is introduced manually which is not convenient or even trivial in case of large (say, $5 \div 7$) number of particles in the final state.

Two possibilities to solve the issue:

- ▶ automatic remapping of the phase space,
- ▶ developing of the integration method less sensitive to the phase space remapping.

The methods described in the report might be a step in the line of the second option.

What Is Markov Chain Monte Carlo?

Markov chain

Let x_i be a point in the phase space of some process. Let the next point x_{i+1} be chosen with the conditional probability $\mathcal{P}\{x_{i+1} \mid x_i\}$, and so on. The important feature is that the probability \mathcal{P} depends on the latest position only but not on the earlier ones. Such a stochastic sequence is called the Markov chain.

MCMC

One can relatively easily construct a Markov chain that mimics a sample from any given probability density function (ProbDF), e.g. SQME of the process of interest. The family of methods to do the job are called Markov chain Monte Carlo (MCMC).

Current Status of MCMC

Although MCMC is quite novel in high energy physics it has been used in statistics for a long time, especially for two recent decades. In Bayesian statistics MCMC is the most popular way to calculate the expectations with respect to very complicated posterior ProbDFs (as elaborate as an artificial neural network weights distribution, for example). So, **MCMC techniques are well tested and understood.**

There is an intense current development, too.

MCMC: pro et contra

- ▶ MCMC can sample points from a quite complicated ProbDF (i.e. it is less sensitive to the sharp “ridges” of the integrand).

But:

- ▶ The points are correlated.
- ▶ MCMC usually needs smart tuning.
- ▶ I mentioned sampling, what is about integration?

In the line of last point the question is

- ▶ whether one can estimate the cross-section given the sample (withstanding the “curse of dimensionality”)
- ▶ and whether the method’s performance is comparable to VEGAS with the phase space remapping applied.

Integration with MCMC (1/5)

In statistics MCMC is typically used to calculate the expectation of the form

$$\mathbb{E}_p[f] = \int d\mathbf{x} f(\mathbf{x}) \frac{p(\mathbf{x})}{\int d\mathbf{y} p(\mathbf{y})} \approx \frac{1}{N} \sum_{i=1}^N f(\mathbf{x}_i),$$

where points $\{\mathbf{x}_i\}$ are samples from (unnormalized) ProbDF $p(\cdot)$ via MCMC. It is not what we want, but one can try

$$\frac{1}{\sigma} = \frac{1}{\int d\mathbf{x} p(\mathbf{x})} = \int d\mathbf{x} \frac{1}{p(\mathbf{x})} \frac{p(\mathbf{x})}{\int d\mathbf{y} p(\mathbf{y})} = \mathbb{E}_p \left[\frac{1}{p} \right]$$

with $p(\cdot)$ standing for the SQME. Since SQME can be close to zero, it is better to use, e.g.,

$$\frac{1}{\sigma + h} = \mathbb{E}_{\tilde{p}} \left[\frac{1}{\tilde{p}} \right], \quad \tilde{p}(\mathbf{x}) = p(\mathbf{x}) + h.$$

Integration with MCMC (2/5)

The method described above provides a computationally feasible way to estimate the integral given the (non-uniform) sample from the SQME. Moreover, under some natural assumptions one can deduce that this is the only possible way to estimate the cross-section by evaluating an expectation of the form

$$\int d\mathbf{x} F[\rho(\mathbf{x})] \frac{\rho(\mathbf{x})}{\int d\mathbf{y} \rho(\mathbf{y})} = \mathbb{E}_\rho [F[\rho]].$$

Given all the other choices of $F(\cdot)$ this expectation cannot be a smooth non-constant function of the integral.

Integration with MCMC (3/5)

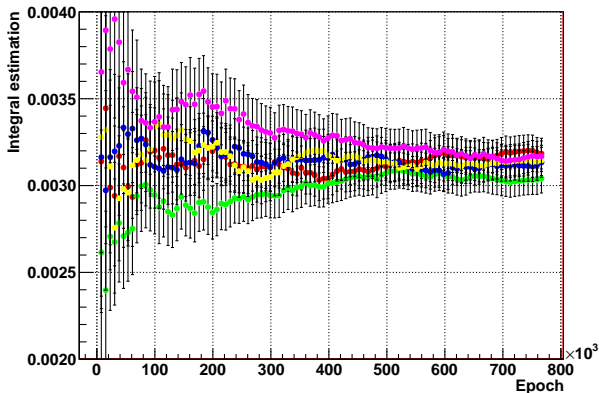


Figure: Integral estimation over several Markov chains in parallel

Integration with MCMC (4/5)

However, one can prove analytically that the method demonstrates larger uncertainty than the simple uniform MC does when applied to Cauchy distribution (and therefore is expected to perform worse given a real SQME). The numerical test also reveals it as being not better than the uniform MC (see the next slide).

Nevertheless, the illustration example is not the only one possibility to estimate the cross-section. One can try to calculate the empirical density of sampled points. Since the density is given by $p(\cdot) / \int d\mathbf{y} p(\mathbf{y})$, it will immediately retrieve the value of the integral. This option will be investigated in the near future.

Integration with MCMC (5/5)

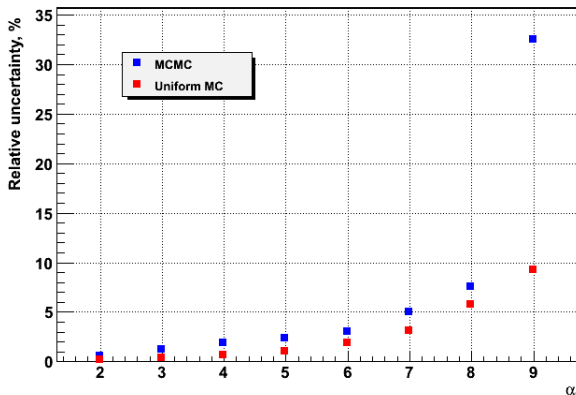


Figure: Errors of MCMC-integration and uniform MC $\left(\int_{-1/2}^{1/2} \frac{dx}{\alpha x^2 + 1} \right)$

Summary and Conclusions

- ▶ MCMC is a well-developed family of methods that can generate a sample from an elaborate ProbDF such as a SQME which is a fascinating feature.
- ▶ To profit from this feature one should design a computationally efficient way to estimate the integral given a sample of the specific form.
- ▶ Thanks to the advantages of MCMC such an algorithm might provide a sufficiently small uncertainty with no (not all) phase space remappings applied.
- ▶ Of course, nothing is guaranteed and the actual performance will depend considerably on the details of the integration technique used.

Backup Slides

Metropolis Algorithm

Metropolis algorithm is the simplest one of all the MCMC family, but nevertheless the most popular. Let $p(\cdot)$ be the (unnormalized) ProbDF to sample from, and $q(\cdot | \cdot)$ be an auxiliary (“proposal”) density that is easy to sample from. The proposal ProbDF is required to be symmetric: $q(\mathbf{x} | \mathbf{y}) = q(\mathbf{y} | \mathbf{x})$; the Gaussian is often used:

$$q(\mathbf{x}' | \mathbf{x}) \sim \exp\{-(\mathbf{x} - \mathbf{x}')^2\}.$$

Algorithm

- ▶ At step i sample a proposal point $\mathbf{x}' \sim q(\cdot | \mathbf{x}_i)$.
- ▶ Accept \mathbf{x}' with probability $\min\{p(\mathbf{x}')/p(\mathbf{x}), 1\}$.
- ▶ If the proposal point is accepted, $\mathbf{x}_{i+1} = \mathbf{x}'$, otherwise the current point is copied in the sample: $\mathbf{x}_{i+1} = \mathbf{x}_i$.

Uncertainty of the Illustration Method

The uncertainty of MCMC method used for illustration is

$$\frac{\delta[I]}{I} = \frac{I+h}{I} \sqrt{(I+h) \int \frac{d\mathbf{x}}{\rho(\mathbf{x})+h} - 1} \sqrt{\frac{\tau}{N}}.$$

Here $I = \int d\mathbf{x} \rho(\mathbf{x})$, N — number of sampled points,

$$\tau = 1 + 2 \sum_{s=1}^{+\infty} \rho(s)$$

is the number of correlated points to give an equivalent of one independent point, $\rho(\cdot)$ — the autocorrelation function:

$$\rho(s) = \frac{1}{\sigma^2} \mathbb{E}_p[(f(\mathbf{x}_i) - \mathbb{E}_p[f])(f(\mathbf{x}_{i+s} - \mathbb{E}_p[f]))],$$

and $f(\cdot)$ is some scalar function of the sampled point.