Nuclear PDF determination using Markov Chain Monte Carlo methods

Nasim Derakhshanian

nCTEQ collaboration

Department of Theoretical Particle Physics ,Institute of Nuclear Pysics PAN, Krakow, Poland

Nasim.derakhshanian@ifj.edu.pl

Introduction

Nuclear parton distribution functions (nPDFs) are essential in the prediction of heavy-ion collisions. The challenging part of nPDF extraction is the uncertainty estimation. The most common approach for this purpose is the Hessian method, which has certain shortcomings, especially in the case of nuclear PDFs. Here I will show a case study for an alternative approach where nPDF uncertainties are estimated using the Markov Chain Monte Carlo (MCMC) methods.

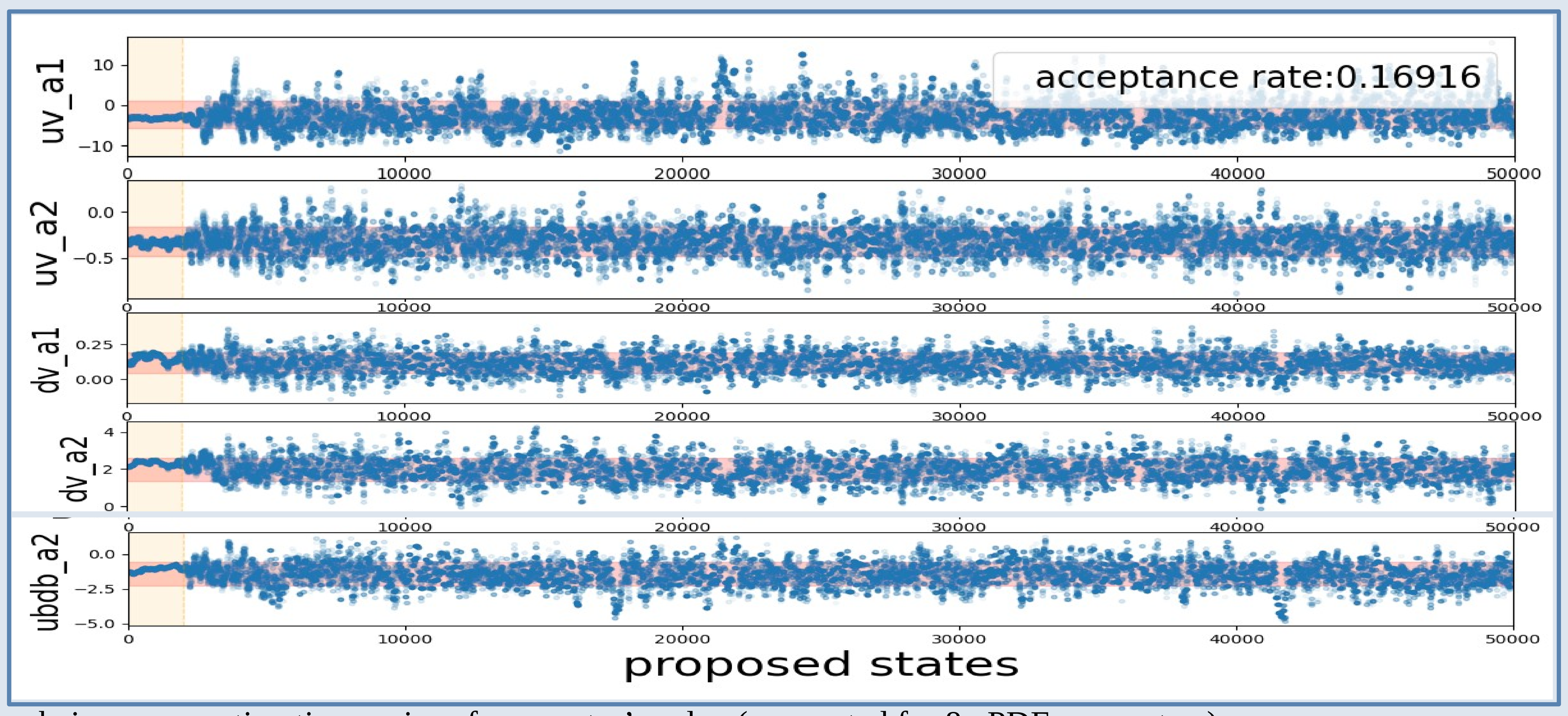
Markov Chain Monte Carlo (MCMC)

MCMC method is a powerful tool to sample complex probability distributions. It is based on building a Markov chain in which each state is made by drawing a sample from a proposal distribution and accepting or rejecting the sample based on an acceptance criterion. The algorithm proceeds by iteratively generating a sequence of samples that converge to the target distribution, allowing for efficient exploration of the parameter space.

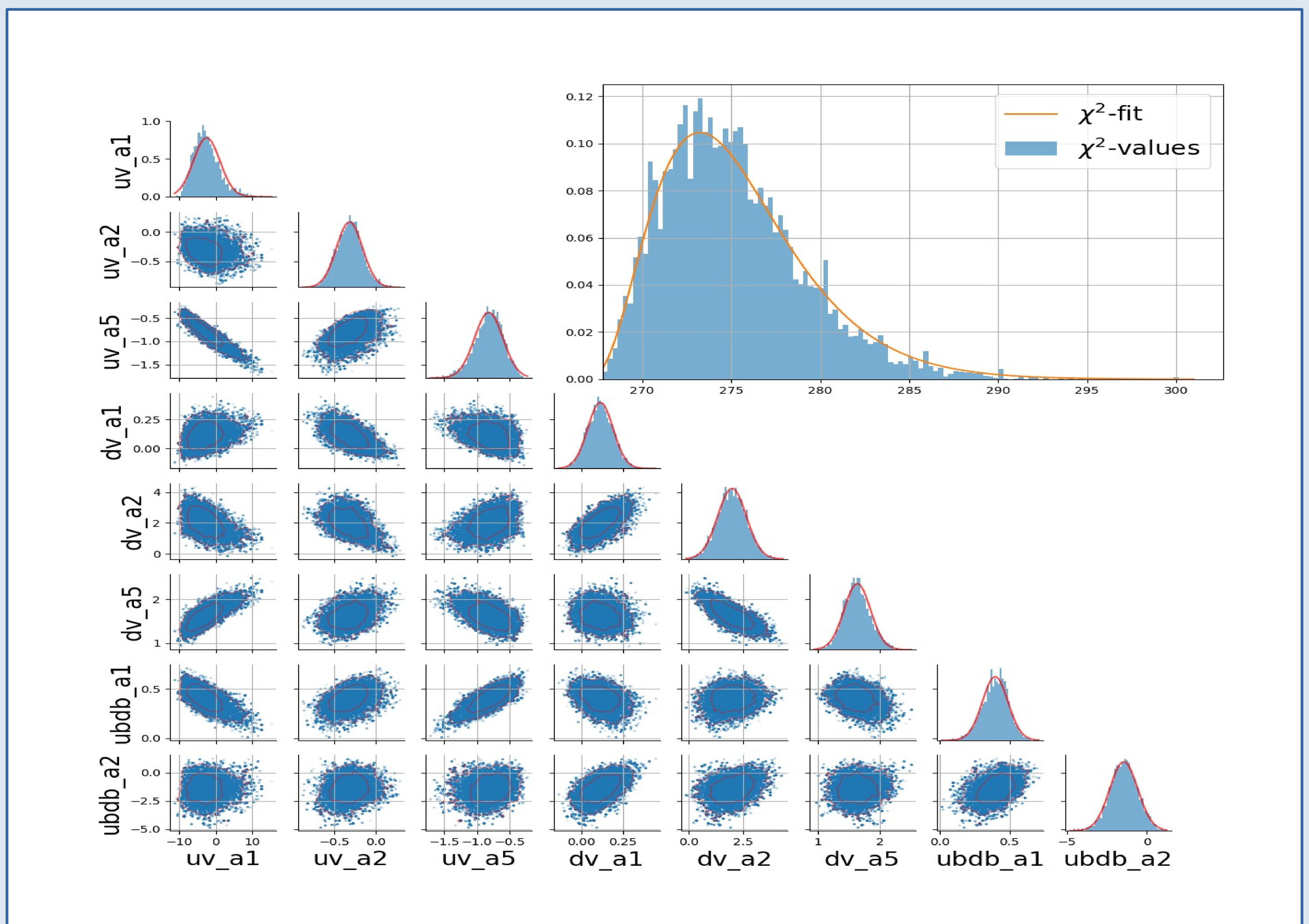
Methodology

The goal is to find the set of nPDF parameters that maximizes the posterior probability distribution given the experimental data.

- Choose the nCTEQ parametrization for the nuclear PDF.
- Construct a likelihood function that incorporates experimental uncertainties and correlations. Using Bayes' theorem, the posterior function is defined in terms of likelihood and prior.
- Use MCMC algorithm (Metropolis-Hastings) to generate a Markov chain of the nPDF parameters that samples from the posterior probability distribution. The algorithm explores the parameter space by generating new sets of parameters based on the previous ones, with a probability that depends on the posterior.
- The uncertainty estimation is done through the analysis of the Markov chain. It can be challenging, particularly when dealing with high-dimensional parameter spaces or complex models. To ensure accurate uncertainty estimates, it is important to carefully evaluate the Markov chain and choose appropriate convergence diagnostics and statistical measures.



chains representing time series of parameter's value (generated for 8 nPDF parameters)



diagonal: histogram of each parameter; off-diagonal: 2D correlation plots between parameters. The inner and outer contours are regions containing respectively 68% and 95% of the probability density

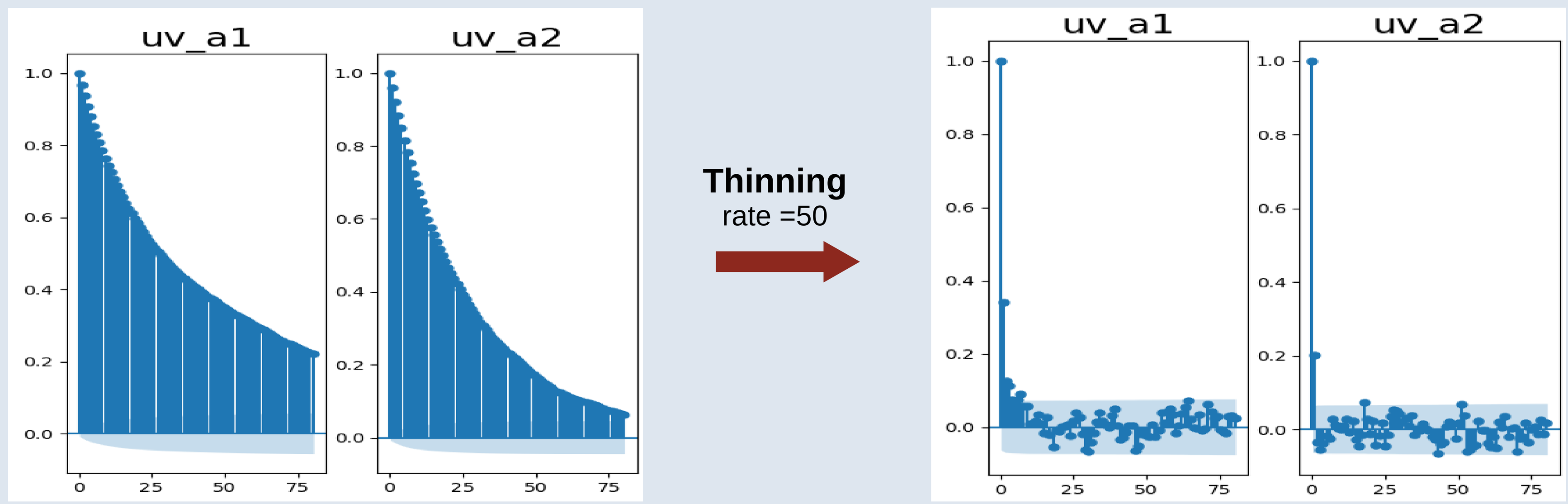
Acknowledgment:
This work was supported by Narodowe Centrum Nauki under grant no.\ 2019/34/E/ST2/00186. I also acknowledge financial support of the Polish National Agency for Academic Exchange, PROM program.

Challenges

Aside from the high computational cost, one of the challenges with using MCMC to find PDFs is dealing with autocorrelation. Since the samples generated in the chain are highly correlated, they do not effectively explore the entire parameter space. This can lead to slow convergence and inefficient sampling, making it difficult to obtain accurate parameter estimates.

Thinning method

Thinning is a method used in MCMC algorithms to reduce autocorrelation in the generated samples. The basic idea of thinning is to keep only every *k*-th sample in the Markov chain and discard the rest. The choice of thinning parameter k should be carefully tuned to balance the reduction in autocorrelation with the loss of efficiency (information).



Autocorrelation function versus time interval

Why Thinning?

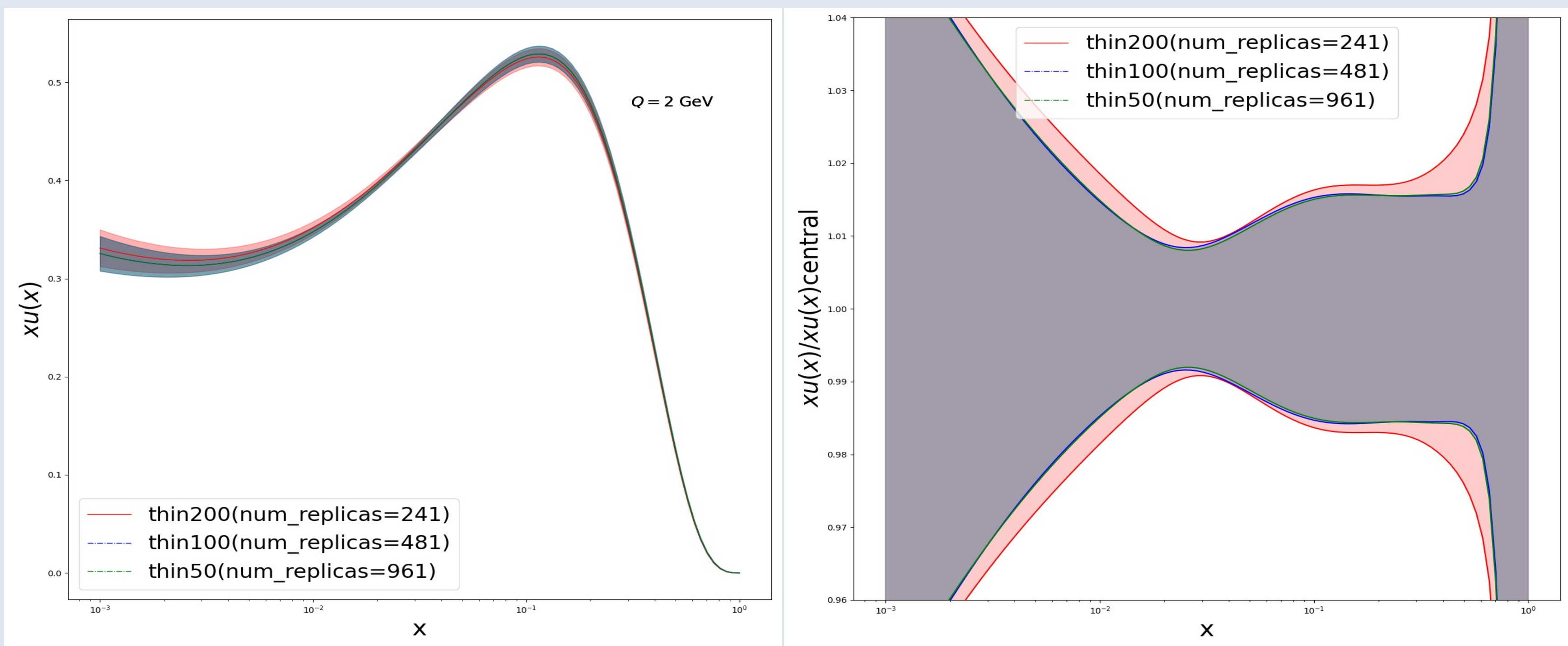
- It provides an **uncorrelated** chain so we can use Monte-Carlo error estimation:

$$\sigma^2_{MCMC} = 2\,\tau_{int}\,\sigma^2_{MC} \quad \longrightarrow \quad \sigma^2_{MC} = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \hat{\mu})^2$$

- We aim to generate a set of PDF grids with corresponding chain units. By limiting the number of points, it makes the grids smaller and user-friendly.

Producing final results - nPDFs

- Generating the Markov Chain
- Analyzing and diagnosing the chain
- Thinning the chain
- Save nPDFs corresponding to each point of the chain in a community standard LHAPDF format. Anyone can use such nPDFs, which make it simple to calculate any nPDF-dependent observables and their uncertainties.



Full-PDF and ratio-PDF for thinned chain of 8 parameters

Conclusion

Despite the challenges, MCMC has become a powerful tool for determining PDFs. It provides a robust and statistically rigorous framework for extracting the PDFs from experimental data.

The promising results that I obtained were for a simplified analysis, and now they need to be scaled up for more parameters and more data.

References

(1) Kovařík, K., et al. "*nCTEQ15: Global analysis of nuclear parton distributions with uncertainties in the CTEQ framework*." Physical Review D 93.8 (2016): 085037.

(2) Risse, P. (2020) *Markov Chain Monte Carlo studies of the Proton Structure (Master's thesis, WWU, Munster, Germany)*