Makov Chain Monte Carlo for PDF uncertainties

A. Kusina
Institute of Nuclear Physics PAN, Krakow, Poland

In collaboration with: P. Risse, N. Derakhshanian, T. Jezo, K. Kovarik

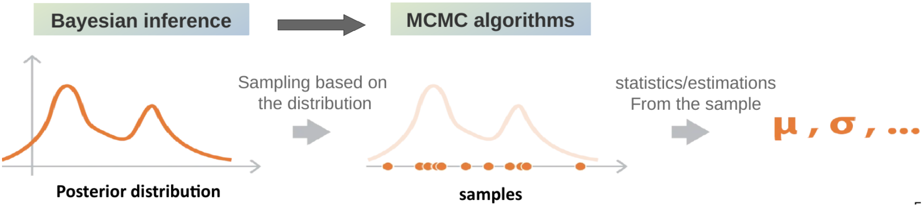
PDF4LHC 2024 meeting
2-3 December 2024, CERN

Work supported by:
Narodowe Centrum Nauki
SONATA BIS grant No 2019/34/E/ST2/00186

nCTEQ
nuclear parton distribution functions

Introduction

- We would like to have access to the underlying distribution of PDFs (PDF parameters) – the **posterior distribtion**
  - this would allow for reliable compution of errors for any PDF dependent quanties
- Posterior distribution too complicated to sample directly
  - need clever way to choose Monte Carlo samples
- Solution: construct the Monte Carlo samples via a Markov chain



Markov Chain Monte Carlo representation of the likelihood

Bayes theorem

$$\pi(\mathbf{c}|\text{data}) = \frac{l(\text{data}|\mathbf{c})\; p(\mathbf{c})}{\mathcal{N}}$$

with: $\pi(\mathbf{c}|\text{data})$ - posterior, $l(\text{data}|\mathbf{c})$ - likelihood, $p(\mathbf{c})$ - prior distribution,
$\mathcal{N} = \int d\mathbf{c}\; l(data|\mathbf{c})\; p(\mathbf{c})$ - normalization

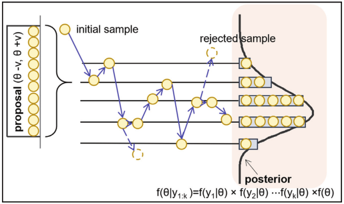
- Likelihood given by the χ^2 function:

$$l(\text{data}|\mathbf{c}) \propto \exp\left(-\frac{1}{2}\chi^2(\mathbf{c}, D)\right)$$

- **Draw random samples** from the posterior function:
  $\pi(\mathbf{c}|\text{data}) \to \{\mathbf{c_1}, \mathbf{c_2}, \ldots, \mathbf{c_n}\}$
- Samples have to **reproduce the expectation value and higher modes**:

$$E\{\mathcal{O}(\mathbf{c})\} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{O}(\mathbf{c}_i)$$

1. Use **normal random walk Metropolis-Hastings**
   - set initial values of the parameters $\mathbf{c}_0$
   - propose set of new parameters: $\tilde{\mathbf{c}}_{i+1}$
     proposal distribution: $q(\tilde{\mathbf{c}}_{i+1}, \mathbf{c}_i) = \mathcal{N}(\mathbf{c}_i, C_0)$
   - compute acceptance probablity:
     $\alpha = \min\left(\frac{\pi(\tilde{\mathbf{c}}_{i+1}|\text{data})}{\pi(\mathbf{c}_i|\text{data})}, 1\right)$
   - if $\mathcal{U}(0,1) < \min(1,\alpha)$: $\mathbf{c}_{i+1} = \tilde{\mathbf{c}}_{i+1}$
     else: $\mathbf{c}_{i+1} = \mathbf{c}_i$



2. After N_0 samples switch to a **self learning proposal distribution**

$$\tilde{\mathbf{c}}_{i+1} \quad \text{proposed from} \quad q(\tilde{\mathbf{c}}_{i+1}, \mathbf{c}_i) = (1-\beta)\mathcal{N}\left(\mathbf{c}_i, \text{scale}\cdot\overline{C}_i\right) + \beta\mathcal{N}(\mathbf{c}_i, C_0)$$

with **self learned** covariance $\overline{\mathbf{C}}_\mathbf{i}$

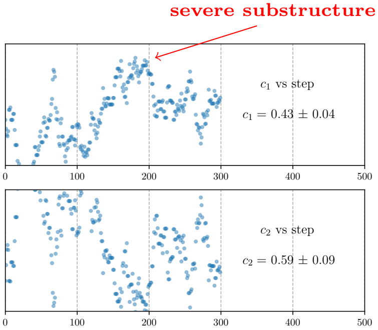
   - $0 \leq \beta \leq 1$ controls the impact of the 'learned' proposal

3. Reset self learned proposal distribution **to boost convergence**
   - this reduces the impact of the starting point

[H. Haario et al.: "An adaptive Metropolis algorithm", *Bernoulli* 7.2 (Apr. 2001)]

Autocorrelation

- Because of the dependence on the previous step MCMC samples are correlated.
- Hence, **standard Monte Carlo error estimates** do not work $\sigma^2_{MC} = \frac{1}{n-1}\Sigma_{k=1}^{n}(X_k - \hat{\mu})^2$
  - these severely underestimate the true uncertainties
- Since every new sample depends on the previous one **the gain in information is reduced**
- This is what is called **autocorrelation**
  - twice the **autocorrelation-time τ** estimates the number of links in the chain **until the next independent sample is drawn**



- Lattice QCD has several methods to deal with this problem
- One example is the **Γ-method**
  - allows to estimate the autocorrelation time, τ_{int}, directly from the chain
  - used to **enlarge error estimates** to eliminate bias: $\sigma^2_{MCMC} = 2\tau_{int}\sigma^2_{MC}$
  - **or filter/thin** the time series to get uncorrelated samples

Monte Carlo errors with less errors.

Ulli Wolff*
Institut für Physik, Humboldt Universität
Newtonstr. 15
12489 Berlin, Germany

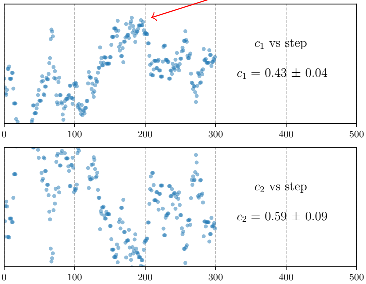
ALPHA Collaboration

Abstract
We explain in detail how to estimate mean values and assess statistical errors for arbitrary functions of elementary observables in Monte Carlo simulations. The method is to estimate and sum the relevant autocorrelation functions, which is argued to produce more certain error estimates than binning techniques and hence to help toward a

[arXiv:hep-lat/0306017]

Filtering based on the **Γ**-method

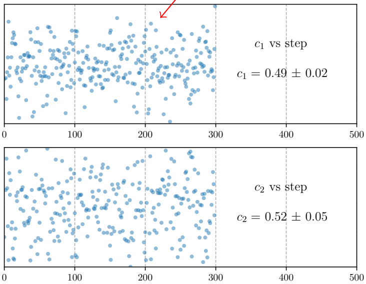
severe substructure



using 300 samples directly

$\Rightarrow$

uniform behaviour

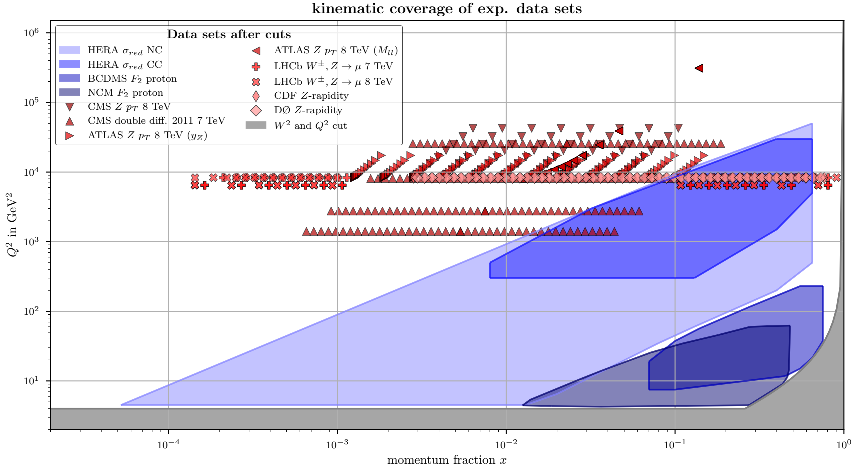


thinning 10^4 samples to a total of 300

Proton PDF analysis

Experimental data

- DIS: 1660 points
  - HERA NC/CC
  - NMC F_2
  - BCDMS F_2
- DY: 324 points
  - CDF & DØ
  - CMS
  - ATLAS
  - LHCb
- Total: 1984 points



Fitting setup

PDF parameters

$$xf_i(x, Q_0) = \mathbf{c_0} x^{\mathbf{c_1}} (1-x)^{\mathbf{c_2}} (1 + \mathbf{c_3}\sqrt{x} + \mathbf{c_4} x)$$

$\mathbf{u_v}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{d_v}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{\overline{u} + \overline{d}}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{s + \overline{s}}$	$\rightarrow$	c_0				
$\mathbf{g}$	$\rightarrow$	c_0	c_1	c_2	c_3	c_4

Total: 15 parameters

Fitting setup

PDF parameters

$$xf_i(x,Q_0) = \mathbf{c_0} x^{\mathbf{c_1}} (1-x)^{\mathbf{c_2}} (1 + \mathbf{c_3}\sqrt{x} + \mathbf{c_4} x)$$

$\mathbf{u_v}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{d_v}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{\bar{u} + \bar{d}}$	$\rightarrow$	c_1	c_2	c_4		
$\mathbf{s + \bar{s}}$	$\rightarrow$	c_0				
$\mathbf{g}$	$\rightarrow$	c_0	c_1	c_2	c_3	c_4

Total: 15 parameters

- Down-valence distribution becomes independent of $\mathbf{c_4}$

$$\lim_{\mathbf{c_4}\to\infty} xd_v(x,Q_0) = \lim_{\mathbf{c_4}\to\infty} c_0 x^{\mathbf{c_1}} (1-x)^{\mathbf{c_2}} [\mathbf{c_4} x]$$
$$= \tilde{\mathbf{c}}_\mathbf{0} x^{\mathbf{c_1}+1} (1-x)^{\mathbf{c_2}}$$

- Need to onstrain $\mathbf{c_4}$ by Uniform Prior:

$$-1000 \leq \mathbf{c_4} \leq 10.000$$

Fitting setup

PDF parameters

$$xf_i(x,Q_0) = \mathbf{c_0} x^{\mathbf{c_1}}(1-x)^{\mathbf{c_2}}(1+\mathbf{c_3}\sqrt{x}+\mathbf{c_4}x)$$

$\mathbf{u_v}$	$\to$	c_1	c_2	c_4		
$\mathbf{d_v}$	$\to$	c_1	c_2	c_4		
$\mathbf{\bar{u}+\bar{d}}$	$\to$	c_1	c_2	c_4		
$\mathbf{s+\bar{s}}$	$\to$	c_0				
$\mathbf{g}$	$\to$	c_0	c_1	c_2	c_3	c_4

Total: 15 parameters

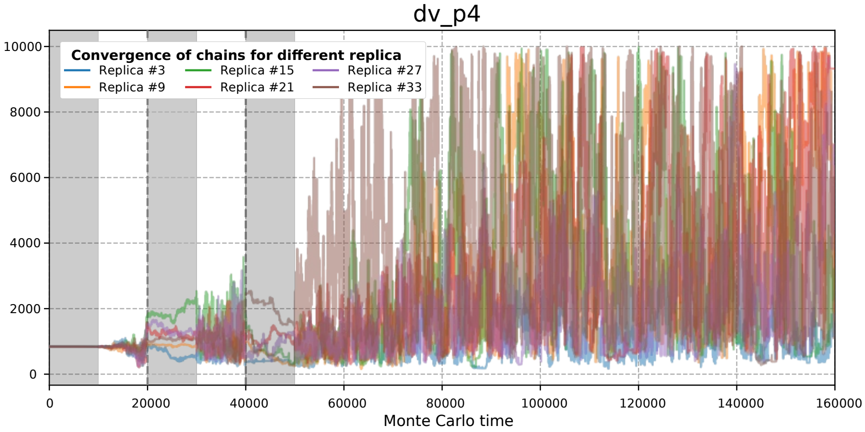
Hyperparameters

- Proposals: Adaptive Metropolis Hastings
- **36 independent chains** with **479k** samples each
  - burn-in phase: 140k samples
  - **Total:** 17 million samples
- removing autocorrelation (thinning rate: 3000) and burn-in:

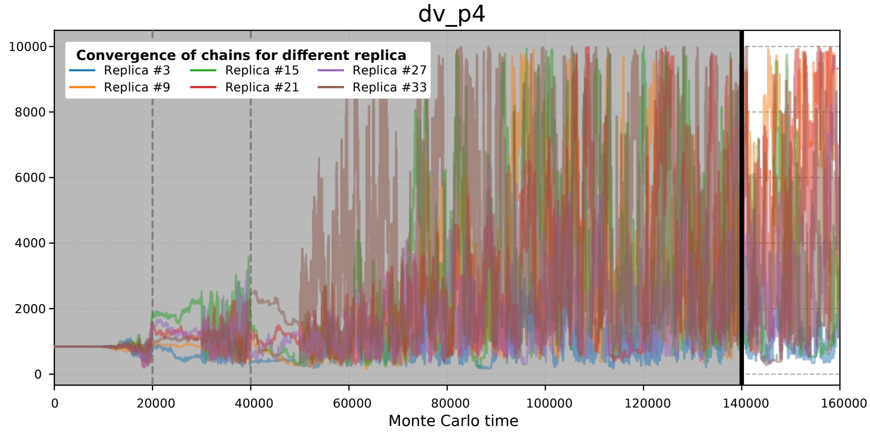
Total: 4068 uncorrelated samples

$$\chi^2/\mathbf{d.o.f.} = \mathbf{2380.25/1969} = \mathbf{1.20}$$

Thermalization



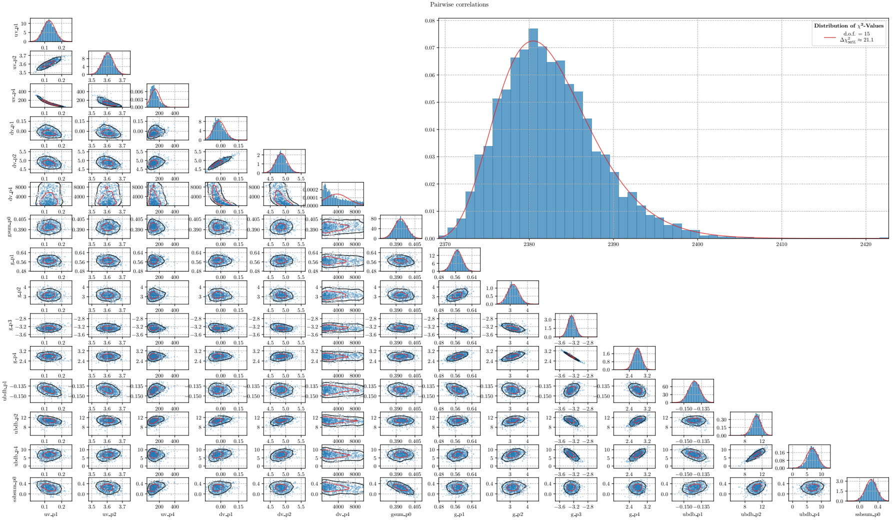
Thermalization



Description of Experimental Data

Data Set	Ref.	Data Points	χ^2/data
DIS			
HERA σ_{red} neutral current	[54]	1039	1.26
HERA σ_{red} charged current	[54]	81	1.08
BCDMS F_2 proton	[135]	339	1.09
NCM F_2 proton	[136]	201	1.54
DIS total		1660	1.25
DY			
CDF Z-rapidity	[137]	28	1.10
DØ Z-rapidity	[138]	28	0.60
ATLAS Z p_T 8 TeV (M_{ll})	[139]	44	1.06
ATLAS Z p_T 8 TeV (y_Z)	[139]	48	0.65
CMS Z p_T 8 TeV	[140]	28	0.46
CMS double diff. 2011 7 TeV	[141]	88	1.02
LHCb $W^{\pm}, Z \to \mu$ 7 TeV	[142]	29	1.07
LHCb $W^{\pm}, Z \to \mu$ 8 TeV	[143]	31	1.18
DY total		324	0.91
Total		1984	**1.20** (per dof)

Combined chain



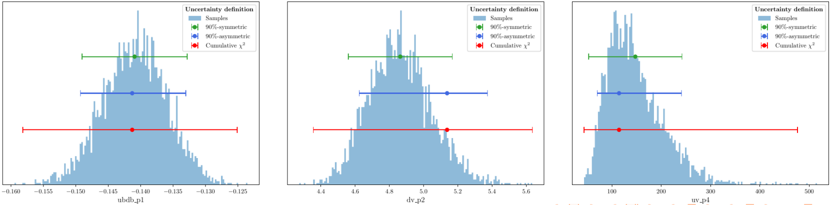
Definition of Uncertainties

From Samples to PDF-Uncertainties

We want to define **confidence interval** for observable $\boldsymbol{\mathcal{O}(c)}$

$$\mathcal{O}_- \leq \mathcal{O} \leq \mathcal{O}_+$$

- α%-symmetric (MCSE):
  - **central value**: mean $\frac{1}{N}\sum_i^N \mathcal{O}_i$
  - **lower (upper) bound**: mean $\pm$ standard deviation
- α%-asymmetric (Percentile/NNPDF-like):
  - **central value**: samle with minimal χ^2 value or 50th percentile
  - **lower (upper) bound**: 16th (84th) percentile of distribution of samples
- Cumulative χ^2:
  - **central value**: sample with minimal χ^2 value
  - **lower (upper) bound**: keep only the samles including lowest α% of samples



Confidence interval for $\mathcal{O}(c)$

$$\mathcal{O}_- \leq \mathcal{O} \leq \mathcal{O}_+$$

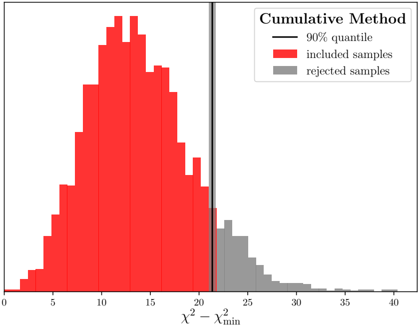
Cumulative χ^2 method

Central: sample with minimal $\chi^2 \to \mathcal{O}_{\chi^2_{min}}$

Lower bound: $\min(\{\mathcal{O}\}_{90\%})$

Upper bound: $\max(\{\mathcal{O}\}_{90\%})$

[A. Putze et al., arXiv: 0808.2437]



Confidence interval for $\mathcal{O}(c)$

$$\mathcal{O}_- \leq \mathcal{O} \leq \mathcal{O}_+$$

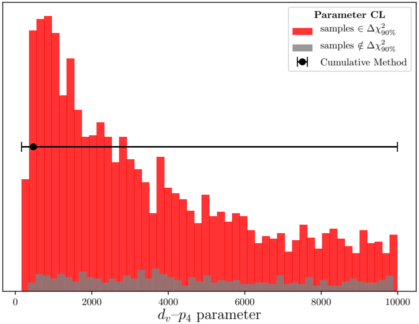
Cumulative χ^2 method

Central: sample with minimal $\chi^2 \rightarrow \mathcal{O}_{\chi^2_{min}}$

Lower bound: $\min(\{\mathcal{O}\}_{90\%})$

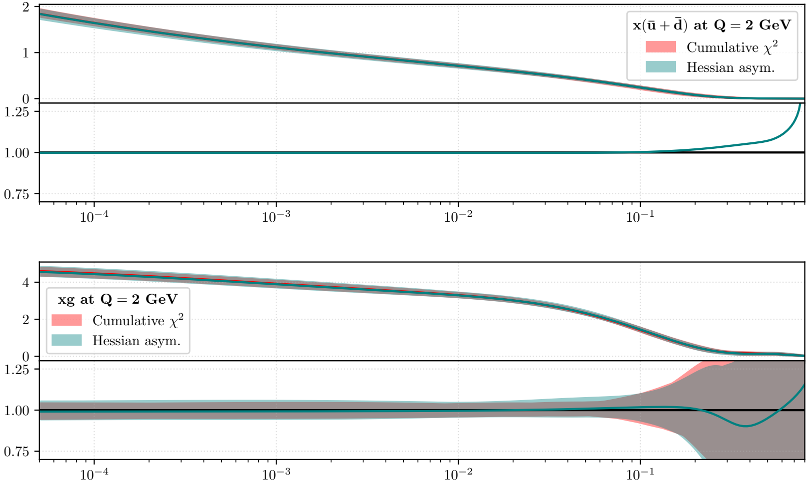
Upper bound: $\max(\{\mathcal{O}\}_{90\%})$

[A. Putze et al., arXiv: 0808.2437]



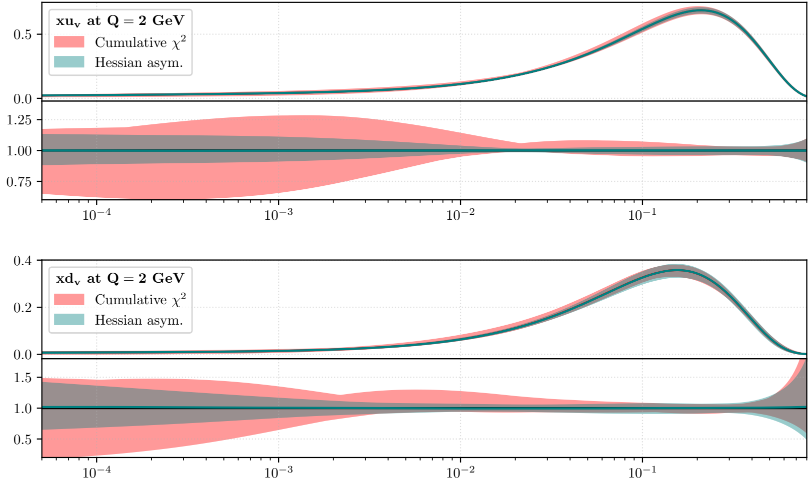
Comparison with Hessian – Gaussian parameters

- **Cumulative χ^2**: $\Delta\chi^2_{90\%} = 22$
- **Hessian Method** $\Delta\chi^2 = 22$



Comparison with Hessian – non-Gaussian parameters

- **Cumulative $\boldsymbol{\chi}^2$**: $\Delta\chi^2_{90\%} = 22$
- **Hessian Method** $\Delta\chi^2 = 22$



Conclusion

Markov Chain Monte Carlo

- Access uncertainties without approximations
- Γ-method to deal with autocorrelation

Proton PDF extraction

- 15 parameters, DIS & DY data
- χ^2/d.o.f. $= 2380.25/1969 = 1.20$
- Generated samples: 17 million
- Result: 4068 uncorrelated samples

Definition of Uncertainties

- Confidence limits using χ^2-values
- Tolerance estim. from χ^2-samples
- comparison with Hessian
  - agreement for Gaussian params.
  - differences for non-Gaussian

Nuclear PDF analysis (only Pb)

Fitting setup

Data & parameterization

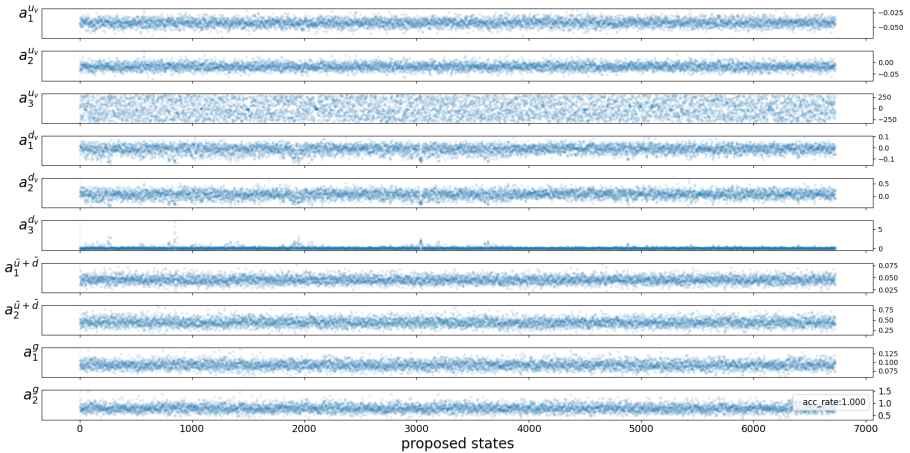
- Data:
  - W and Z boson from pPb LHC
  - heavy quark(onia) from pPb LHC
  - Neutrino DIS
  - Total no data points: **1488**
- Parametrization:
  - $xf_i(x, Q_0) = \mathbf{c_0} x^{\mathbf{c_1}} (1-x)^{\mathbf{c_2}} (1 + \mathbf{c_3}\sqrt{x} + \mathbf{c_4} x)$
  - **10 open parameters** (6 valence, 2 gluon, 2 sea)
- Theory prediction @ NLO
  - HQ with Crystal Ball (data driven)

Hyperparameters

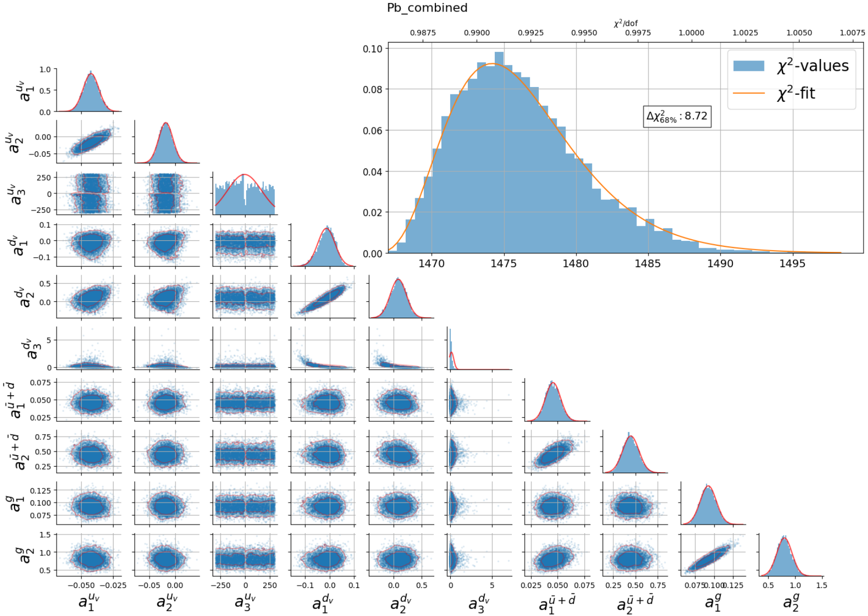
- Proposals: Adaptive Metropolis Hastings
- **9 independent chains** with total of **3.8** million samples
  - burn-in phase: $\approx 60k - 80k$ for each chain
  - Total: **6729 uncorrelated samples**

Combined chain

- After removing autocorrelation all 9 chains are thinned and combined
- No independent sample: $N_{tot} = 6729$.



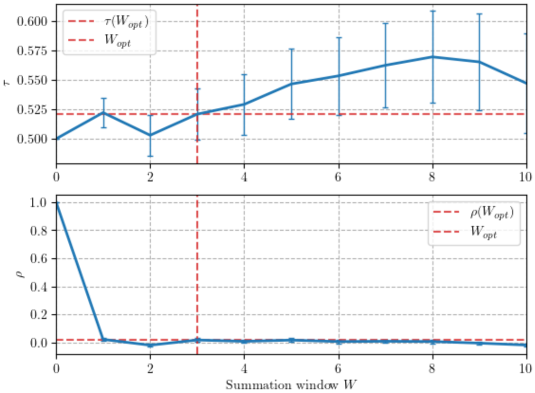
Combined chain



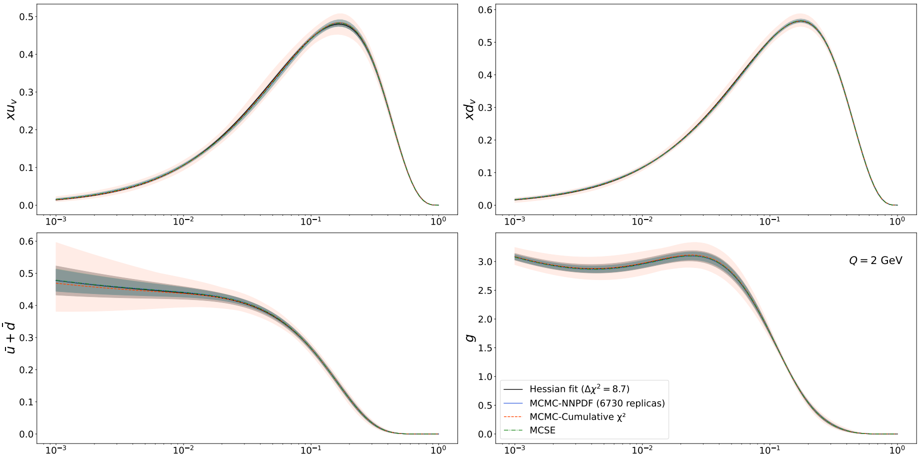
Combined chain: autocorrelation

- Atocorrelation-time τ for te combined chain (using the Γ-method)

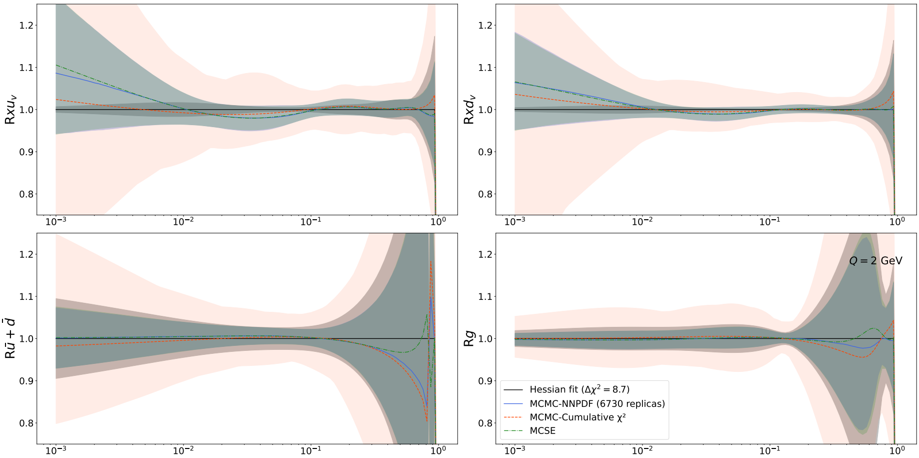
$$\tau \approx \frac{1}{2} + \sum_{t=1}^{\infty} \rho(t)$$



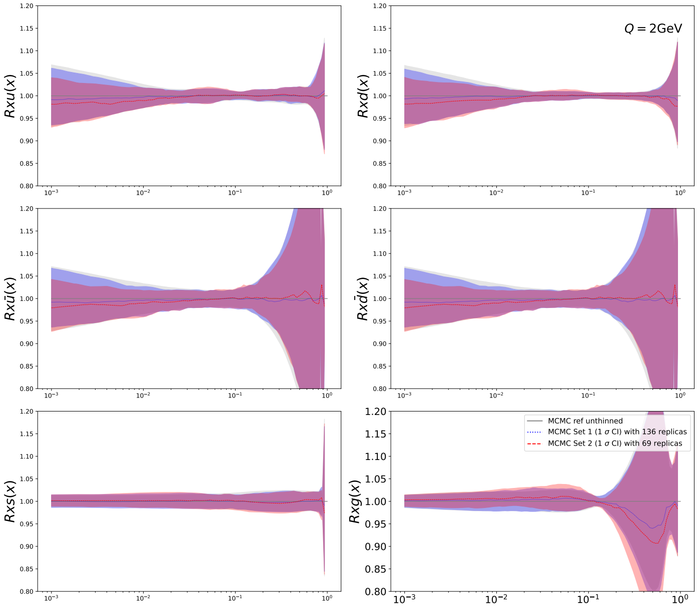
Pb PDFs: Hessian vs MCMC



Pb PDF: Hessian vs MCMC



Pb PDF: Different thinning rates



Conclusion

Markov Chain Monte Carlo

- Access uncertainties without approximations
- Γ-method to deal with autocorrelation

Proton PDF extraction

- 15 parameters, DIS & DY data
- χ^2/d.o.f. $= 2380.25/1969 = 1.20$
- Generated samples: 17 million
- Result: 4068 uncorrelated samples

Nuclear (Pb) PDF extraction

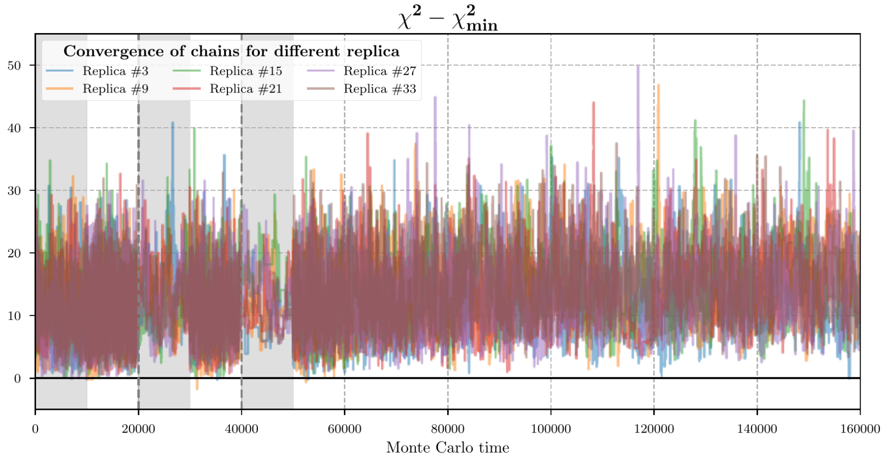
- 10 parameters, CC DIS, W/Z, HQ data
- Generated samples: 3.8 million
- Result: 6729 uncorrelated samples

Definition of Uncertainties

- Confidence limits using χ^2-values
- Tolerance estim. from χ^2-samples
- comparison with Hessian
  - agreement for Gaussian params.
  - differences for non-Gaussian

BACKUP SLIDES

Time series


$\chi^2 - \chi^2_{min}$
Convergence of chains for different replica
Replica #3
Replica #9
Replica #15
Replica #21
Replica #27
Replica #33
0
10
20
30
40
50
0
20000
40000
60000
80000
100000
120000
140000
160000
Monte Carlo time

Thermalization vs. autocorrelation