A Linear Model for Bayesian PDF Inference

Mark N. Costantini

Chiefa et. al, 2501.10359, JHEP

Costantini et. al, 2404.10056, JHEP

Costantini et. al, 2510.03391, Eur. Phys. J.C

Costantini et. al, 2507.16913, JHEP

Nikhef, April 2026





Colibri

Artwork by: @qftoons





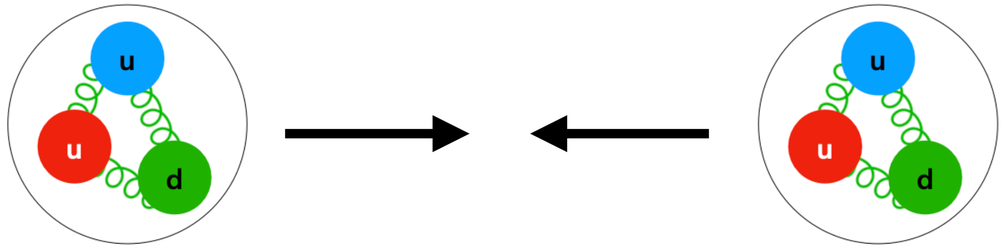
Introduction

Large Hadron Collider

Introduction

World's largest and most powerful particle accelerator - at CERN, Geneva

Two high-energy proton beams travel at close the speed of light before being made to collide



Protons are **composite objects**: bound states of quarks and gluons held together by the strong force

Their internal structure cannot be computed from first principles - it must be inferred from the data

Parton Distribution Functions

The Factorisation Theorem

To make theoretical predictions at the LHC, we use the **factorisation theorem**

Hard-scattering cross sections separate into two independent pieces:

- a **perturbatively calculable** one: the partonic cross section $\hat{\sigma}$

- a non-perturbative part: the parton distribution functions (PDFs), $f_q(x, Q^2)$

$f_q(x, Q^2)$ ~ probability density of finding parton q carrying momentum fraction x at scale Q^2

PDFs are universal: determined once, used for any process

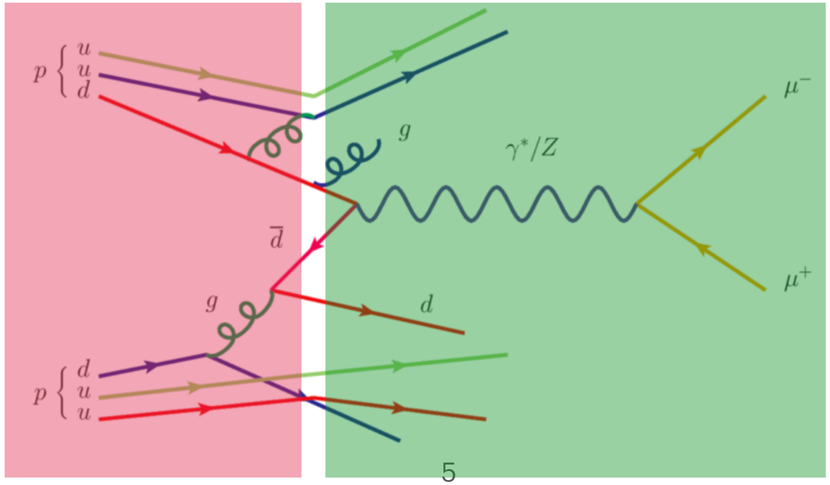
Parton Distribution Functions

The Factorisation Theorem

Observables T at the LHC can be written as:

$$T = \sum_{\text{partons}} \int dx_1 dx_2 f_{q1}(x_1, Q^2) f_{q2}(x_2, Q^2) \hat{\sigma}(x_1, x_2, Q^2)$$

Needs to be determined from data (Universal)



computed in perturbation theory (process dependent)

Parton Distribution Functions

Determining $f(x, Q^2)$ from data

PDFs are functions of 2 variables: $f(x, Q^2)$ (one for each of the active flavours)

Q^2: Energy scale of the observable

The dependence on the scale Q is "controlled" by so-called DGLAP evolution equations

$$Q^2 \frac{\partial f}{\partial Q^2} = \frac{\alpha_S}{2\pi}(P \otimes f)$$

Parton Distribution Functions

Determining $f(x, Q^2)$ from data

PDFs are functions of 2 variables: $f(x, Q^2)$

x: the fraction of momentum carried by the parton

$\rightarrow$ supervised regression problem

Roughly:

given some data $\mathbf{D}$, determine f s.t. $\mathbf{T}[f]$ is as close as possible to $\mathbf{D}$ according to some metric

$$\chi^2 = (\mathbf{D} - \mathbf{T}[f])^T C_D^{-1} (\mathbf{D} - \mathbf{T}[f])$$

Parton Distribution Functions

Different parametrisations

Different groups adopt different functional forms

MSHT

$$x\,f_k(x,Q_0^2) = A_k\,x^{\delta_k}\,(1-x)^{\eta_k}\left(1+\sum_i a_i T_i(y(x))\right)$$

NNPDF

$$x\,f_k(x,Q_0^2) = A_k\,x^{1-\alpha_k}\,(1-x)^{\beta_k}\,\mathrm{NN}_x(x;\theta)$$

Sum Rules

Kinematic limit

Parton Distribution Functions

Different input datasets

Different collaborations include different experimental datasets in their fits

Data set	Ref.	NNPDF3.1	NNPDF4.0	ABMP16	CT18	MSHT20
ATLAS W,Z 7 TeV ($\mathcal{L} = 35$ pb^{-1})	[53]	✓	✓	✓	✓	✓
ATLAS W,Z 7 TeV ($\mathcal{L} = 4.6$ fb^{-1})	[54]	✓	✓	✗	(✓)	✓
ATLAS low-mass DY 7 TeV	[55]	✓	✓	✗	(✓)	✗
ATLAS high-mass DY 7 TeV	[56]	✓	✓	✗	(✓)	✓
ATLAS W 8 TeV	[81]	✗	(✓)	✗	✗	✓
ATLAS DY 2D 8 TeV	[80]	✗	✓	✗	✗	✓
ATLAS high-mass DY 2D 8 TeV	[79]	✗	✓	✗	(✓)	✓
ATLAS $\sigma_{W,Z}$ 13 TeV	[83]	✗	✓	✓	✗	✗
ATLAS W+jet 8 TeV	[95]	✗	✓	✗	✗	✓
ATLAS Z p_T 7 TeV	[274]	(✓)	✗	✗	(✓)	✗
ATLAS Z p_T 8 TeV	[65]	✓	✓	✗	✓	✓
ATLAS $W + c$ 7 TeV	[85]	✗	✓	✗	(✓)	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	[67]	✓	✓	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 7, 8 TeV	[275–280]	✗	✗	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV ($\mathcal{L} = 3.2$ fb^{-1})	[68]	✓	✗	✓	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ 13 TeV ($\mathcal{L} = 139$ fb^{-1})	[136]	✗	✓	✗	✗	✗
ATLAS $\sigma_{t\bar{t}}^{\rm tot}$ and Z ratios	[281]	✗	✗	✗	✗	(✓)
ATLAS $t\bar{t}$ lepton+jets 8 TeV	[69]	✓	✓	✗	✓	✓
ATLAS $t\bar{t}$ dilepton 8 TeV	[91]	✗	✓	✗	✗	✓
ATLAS single-inclusive jets 7 TeV, R=0.6	[75]	✓	(✓)	✗	✓	✓
ATLAS single-inclusive jets 8 TeV, R=0.6	[88]	✗	✓	✗	✗	✗
ATLAS dijets 7 TeV, R=0.6	[148]	✗	✓	✗	✗	✗
ATLAS direct photon production 8 TeV	[102]	✗	(✓)	✗	✗	✗
ATLAS direct photon production 13 TeV	[103]	✗	✓	✗	✗	✗

E.g ATLAS dijets 7 TeV is only included by NNPDF

Parton Distribution Functions

Different uncertainty quantification methods

Groups also differ in how they propagate experimental uncertainties to PDFs

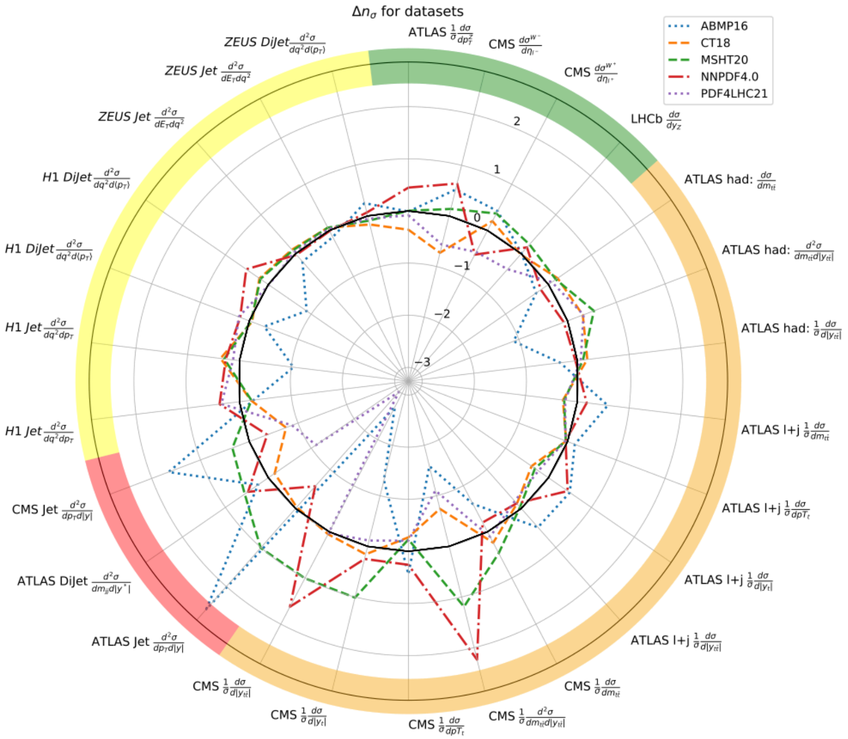
NNPDF: Monte Carlo Replica Method, approximate Bayesian inference with pseudo-data resampling

MSHT, CTEQ: Hessian method with tolerance

Parton Distribution Functions

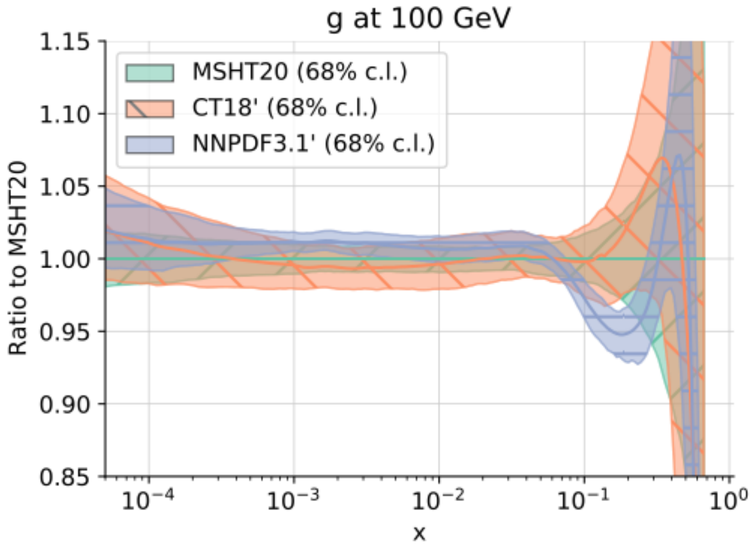
How large are the differences between PDF sets?

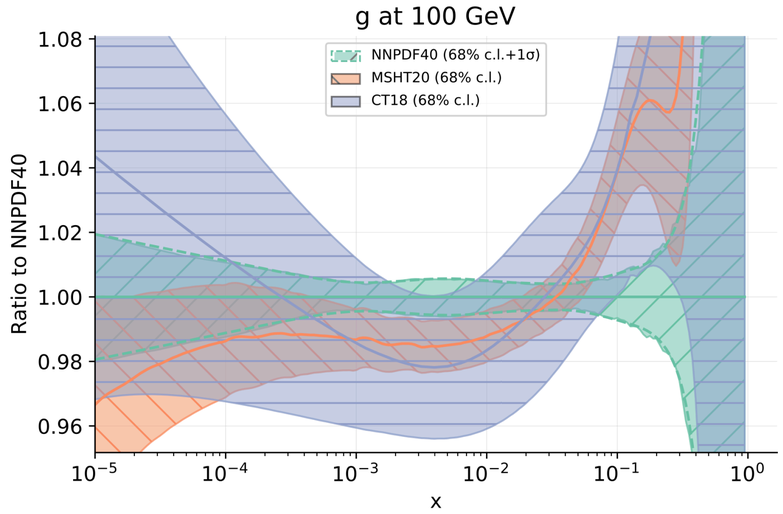
Several studies have been performed to assess how strong differences between PDF sets are



2501.10359, JHEP

PDF4LHC21





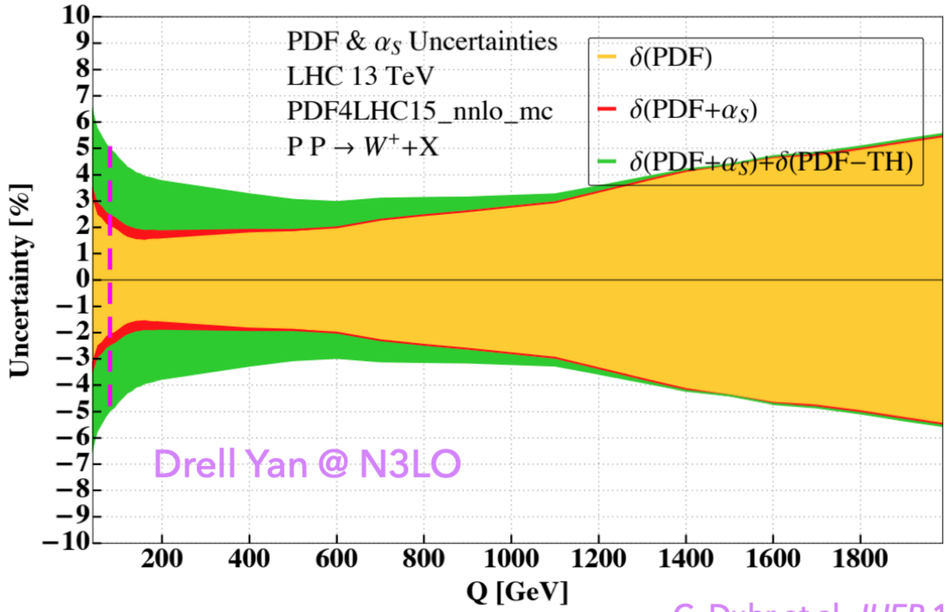
2203.05506,J.Phy.G

PDF Uncertainties

Parton Distribution Functions

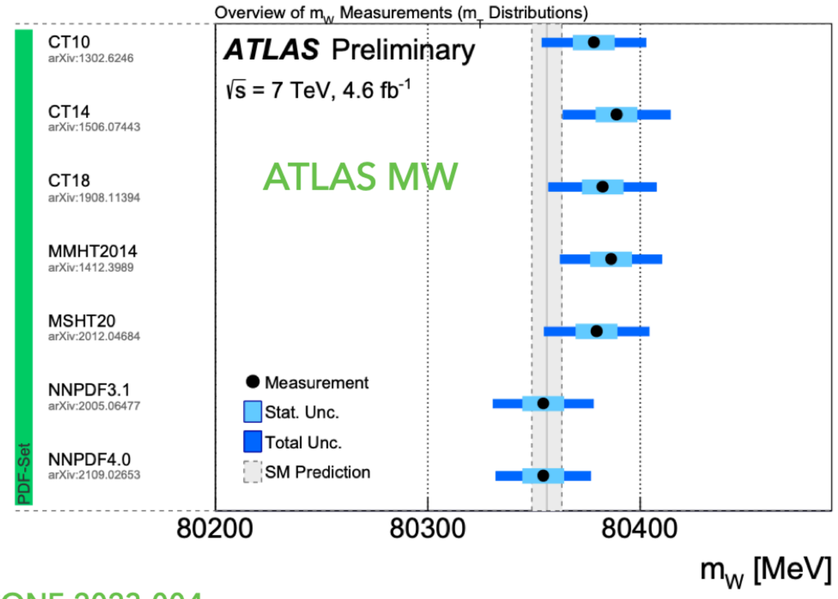
Crucial for precision physics

#1: Theory uncertainty of SM predictions



C. Duhr et al, *JHEP* 11 (2020) 143

#2: Determination of SM parameters



ATLAS-CONF-2023-004

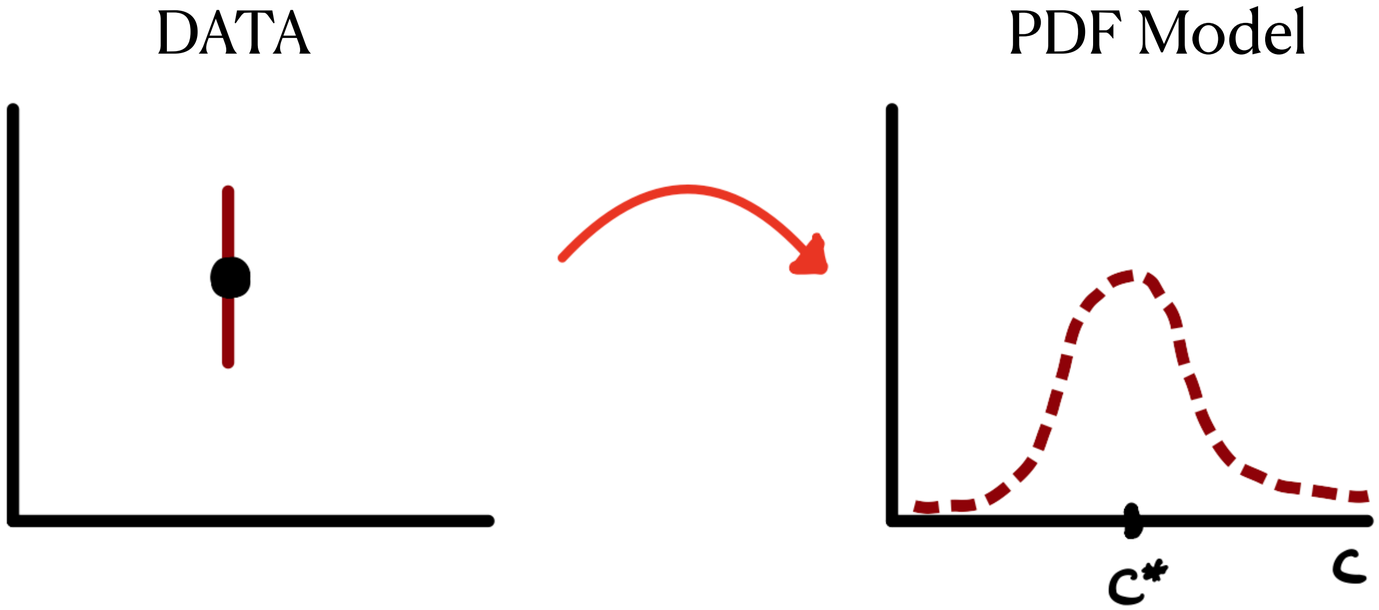
Uncertainty propagation

Setup of the problem

X : observed data

$X \sim \mathbb{P}$: Data generating process

$t[\boldsymbol{c}]$: Modelling of the data



What is a plausible range of values for the parameter(s) $\boldsymbol{c}$?

2404.10056, JHEP

Monte Carlo replica method

Single-point parametrical bootstrap

Use the knowledge of $\mathbb{P}$ to generate N (pseudo) data samples

$$X_1, \ldots, X_N \sim \mathbb{P}$$

for each of the samples we compute an estimate of the parameter

$$\{\boldsymbol{c}_1^*, \ldots, \boldsymbol{c}_N^*\}$$

e.g. $\boldsymbol{c}_i^* = \text{argmin}_{\boldsymbol{c}} \log(L(X_i|\boldsymbol{c}))$

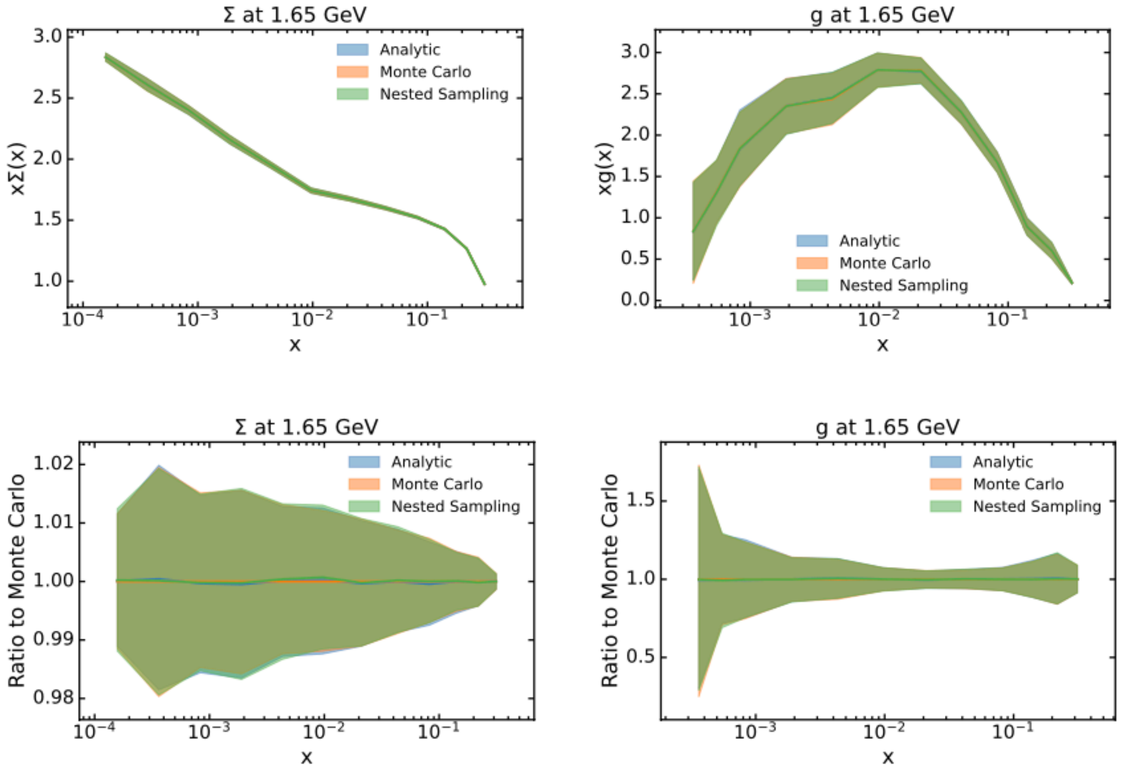
use this collection of samples as an approximation of the posterior

Problem: generally we don't know $\mathbb{P}$, we only have one sample from it

2404.10056, JHEP

PDF toy model fit

DIS (linear) data

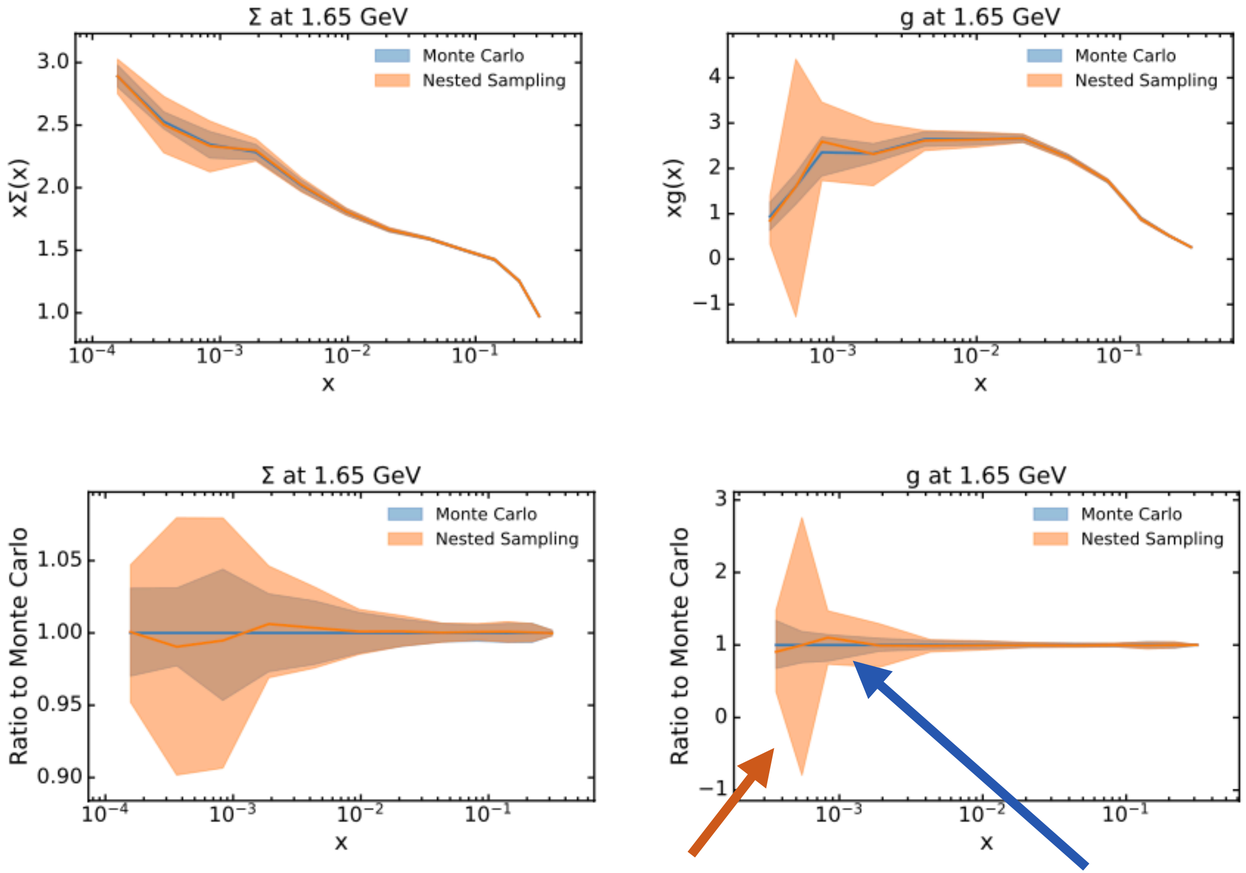


DIS data ~ 3000 datapoints

2404.10056, JHEP

PDF toy model fit

Hadronic-only fit



Hadronic data ~ 1000 datapoints

Nested Sampling

Monte Carlo

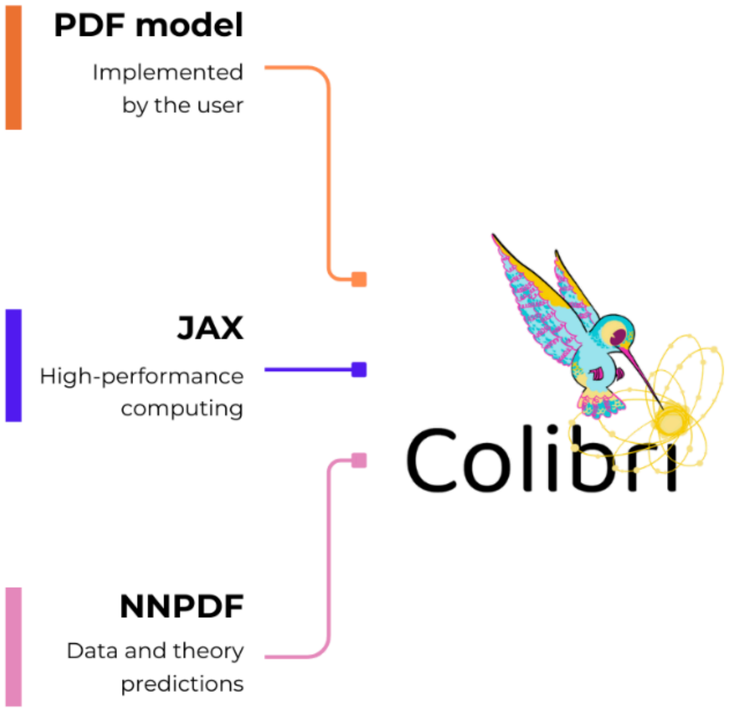
2404.10056, JHEP

Colibri: a PDF fitting framework

Colibri

Comparing inference strategies on equal footing

Colibri is an open-source PDF fitting tool that allows the user to control what **model** to fit and with which **inference method** to use

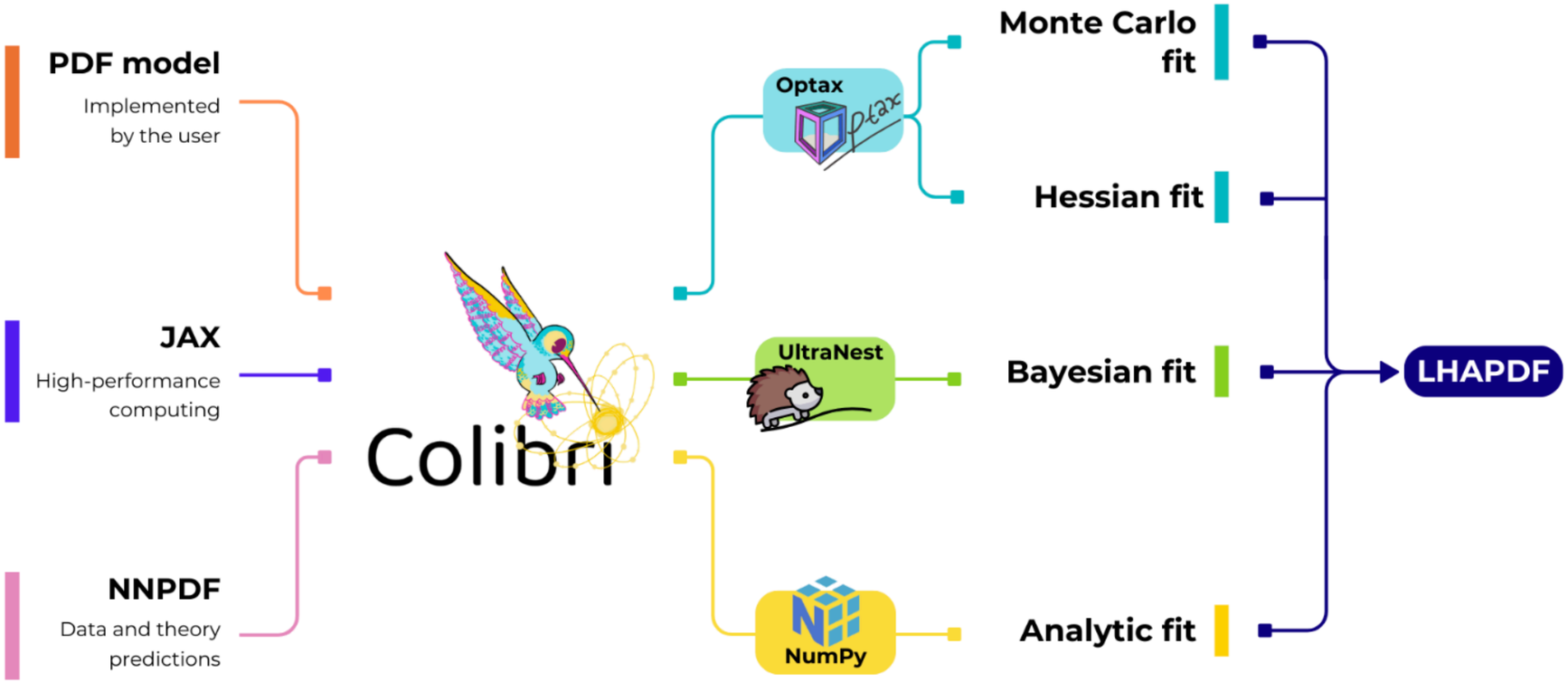


2510.03391, Eur. Phys. J.C

Colibri

Comparing inference strategies on equal footing

Colibri is an open-source PDF fitting tool that allows the user to control what **model** to fit and with which **inference method** to use



2510.03391, Eur. Phys. J.C

A linear PDF model

PDF parametrisation(s)

What makes a good PDF parametrisation?

1. Should respect known theoretical constraints such as **small- and large-x scaling** and **sum rules**

$$f(x) \sim A\, x^{1-\alpha}(1-x)^{\beta}$$

2. It should be flexible enough to explore the space of candidate PDFs amongst $C^1[0,1]$

3. It should be straightforward to fit the model parameters

2507.16913, JHEP

Why a linear parametrisation?

Linear model + linear theory = exact Bayesian inference

The model

The PDF is written as a linear expansion in basis functions:

$$f_w(x,Q^2) = \varphi_0(x,Q^2) + \sum_{i=1}^{N} w_i\,\varphi_i(x,Q^2)$$

The theory prediction (DIS)

For DIS observables, the FK-table convolution is also linear in w

$$T(w) = FK \otimes \varphi_0 + \sum_i w_i (FK \otimes \varphi_i) \equiv X\mathbf{w}$$

The (DIS) data model

$$\mathbf{D} = X\mathbf{w} + \varepsilon, \quad \varepsilon \sim N(0,\Sigma)$$

Analytic posterior

$$p(\mathbf{w} \mid \mathbf{D}) \propto N\left(\hat{\mathbf{w}}, \Lambda^{-1}\right), \qquad \Lambda = X^\top \Sigma^{-1} X, \qquad \hat{\mathbf{w}} = \Lambda^{-1} X^\top \Sigma^{-1} \mathbf{D}$$

Next: how to choose the basis functions φ_i **optimally**?

2507.16913, JHEP

Choosing the basis functions φ_i

Efficiency matters: the basis should be tailored to PDFs

Any complete set works in the large-N limit: $\varphi_i(x) \in$ Chebyshev, Fourier, Legendre, ...

Problem: these bases are designed for generic smooth functions, not for PDF shapes (large and small x suppression)

Goal: choose a basis that is designed around the actual PDF functional space

We need a large, realistic sample of candidate PDFs to learn the basis from.

2507.16913, JHEP

The input sample

Data-unbiased, flexible universal approximants

Naïve choice: existing PDF sets	Better choice: randomly initialised NN's
CT18, MSHT20, NNPDF4.0	NNPDF4.0 architecture, random weights
Already trained on data	Never seen data
-> reduced generalisation capability	-> Covers the full PDF functional space

Each sample has the form:

$$xg_m^j(x, Q_0) = A_m^j \, x^{1-\alpha_m^j}(1-x)^{\beta_m^j} \, \mathrm{NN}_m(x;\theta)$$

Generate $M \sim 20'000$ initialisations, A_m^j fixed by enforcing sum rules on each sample

2507.16913, JHEP

Proper Orthogonal Decomposition

SVD of the ensemble → orthogonal modes ordered by importance

The algorithm

Stack the mean-centred sample into a matrix

$$\tilde{X} = [\tilde{g}_1, \ldots, \tilde{g}_M], \quad \tilde{g}_m = g_m - \varphi_0$$

Perform SVD: $\tilde{X} = USV^T$

The POD basis vectors are the columns of U, ordered by singular value (most to least important)

Theory constraints

Theory constraints are inherited automatically

If all M samples satisfy the sum rules, so does every POD basis vector

2507.16913, JHEP

Tests and applications

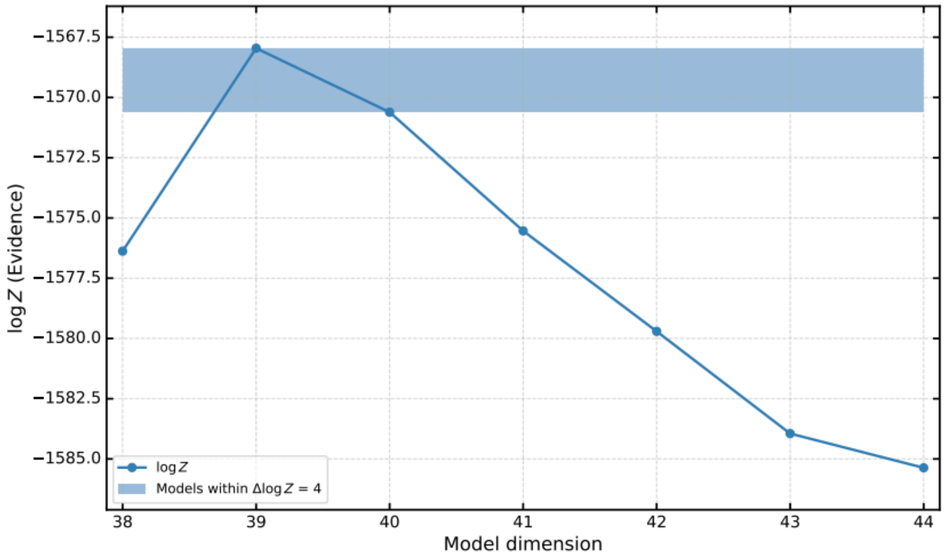
Closure tests

Systematic validation with synthetic data

Choose an **underlying law** f_{in} (a specific POD linear combination with $N_{true} = 40$ active components) and use it to generate pseudo data

$$f_{in} = \varphi_0 + \sum_{i=1}^{40} w_i^{in} \varphi_i, \qquad D_{in} = T[f_{in}] + \varepsilon$$

Verify that the evidence selects the correct model complexity

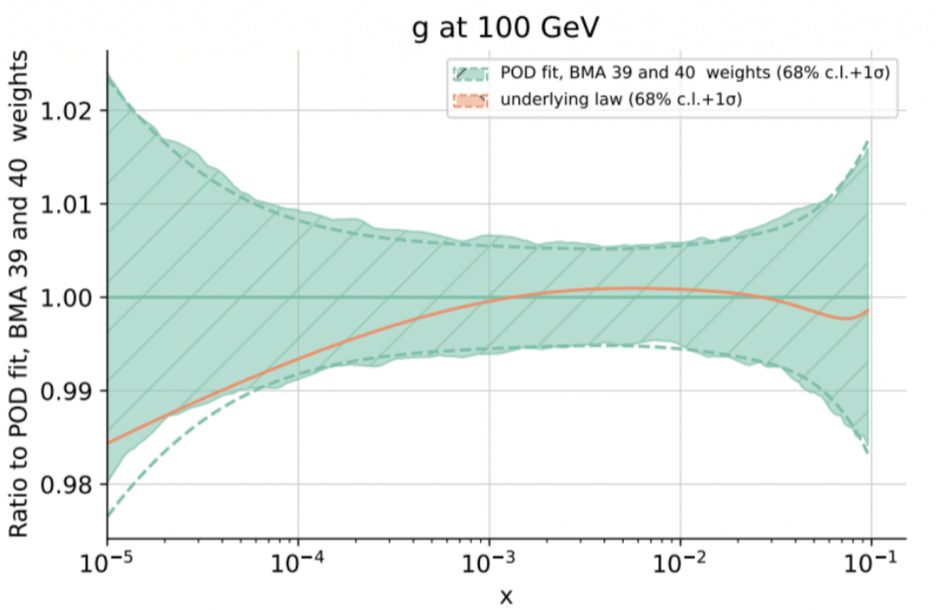


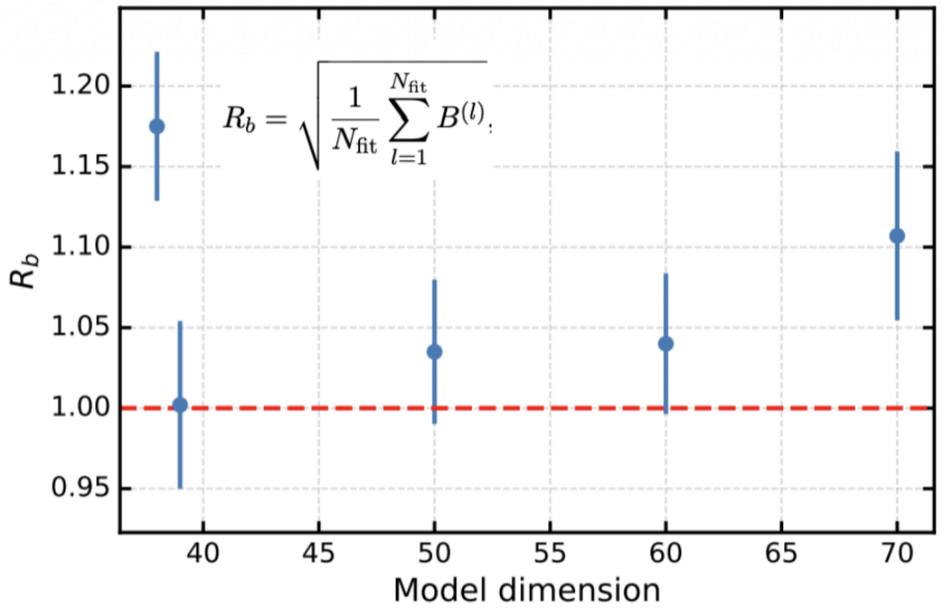
Model Complexity	χ^2/N_{data}	$p(\mathcal{M}_k\|D)$
39	1.033	0.934
40	1.032	0.066

Dataset	N_{data}	Experiment	Model
Numerical fit			
NMC F_2^d/F_2^p [55]	121	Deuteron	Ratio
SLAC F_2^p [56]	33	Deuteron	Linear
SLAC F_2^d [56]	34	Deuteron	Linear
BCDMS F_2^p [57]	333	Deuteron	Linear
BCDMS F_2^d [57]	248	Deuteron	Linear
Analytical fit			
(*) NMC $\sigma^{\text{NC,p}}$ [58]	204	NMC	Linear
CHORUS σ^{ν}_{CC} [59]	416	Nuclear	Linear
CHORUS $\sigma^{\bar{\nu}}_{CC}$ [59]	416	Nuclear	Linear
NuTeV dimuon σ^{ν}_{CC} [60]	39	Nuclear	Linear
(*) NuTeV dimuon $\sigma^{\bar{\nu}}_{CC}$ [60]	37	Nuclear	Linear
HERA I+II $\sigma^{e^+p}_{CC}$ [61]	39	HERA	Linear
HERA I+II charm σ_{NC} [62]	37	HERA	Linear
HERA I+II $\sigma^{e^-p}_{NC}$ (320 GeV)	159	HERA	Linear
(*) HERA I+II $\sigma^{e^+p}_{NC}$ (460 GeV)	204	HERA	Linear
HERA I+II $\sigma^{e^+p}_{NC}$ (575 GeV)	254	HERA	Linear
HERA I+II $\sigma^{e^+p}_{NC}$ (820 GeV)	70	HERA	Linear
HERA I+II $\sigma^{e^+p}_{NC}$ (920 GeV)	377	HERA	Linear
(*) HERA I+II $\sigma^{e^-p}_{CC}$ [61]	42	HERA	Linear
HERA I+II bottom σ_{NC} [62]	26	HERA	Linear
Total	3089		

Closure tests

Uncertainty bands are statistically faithful



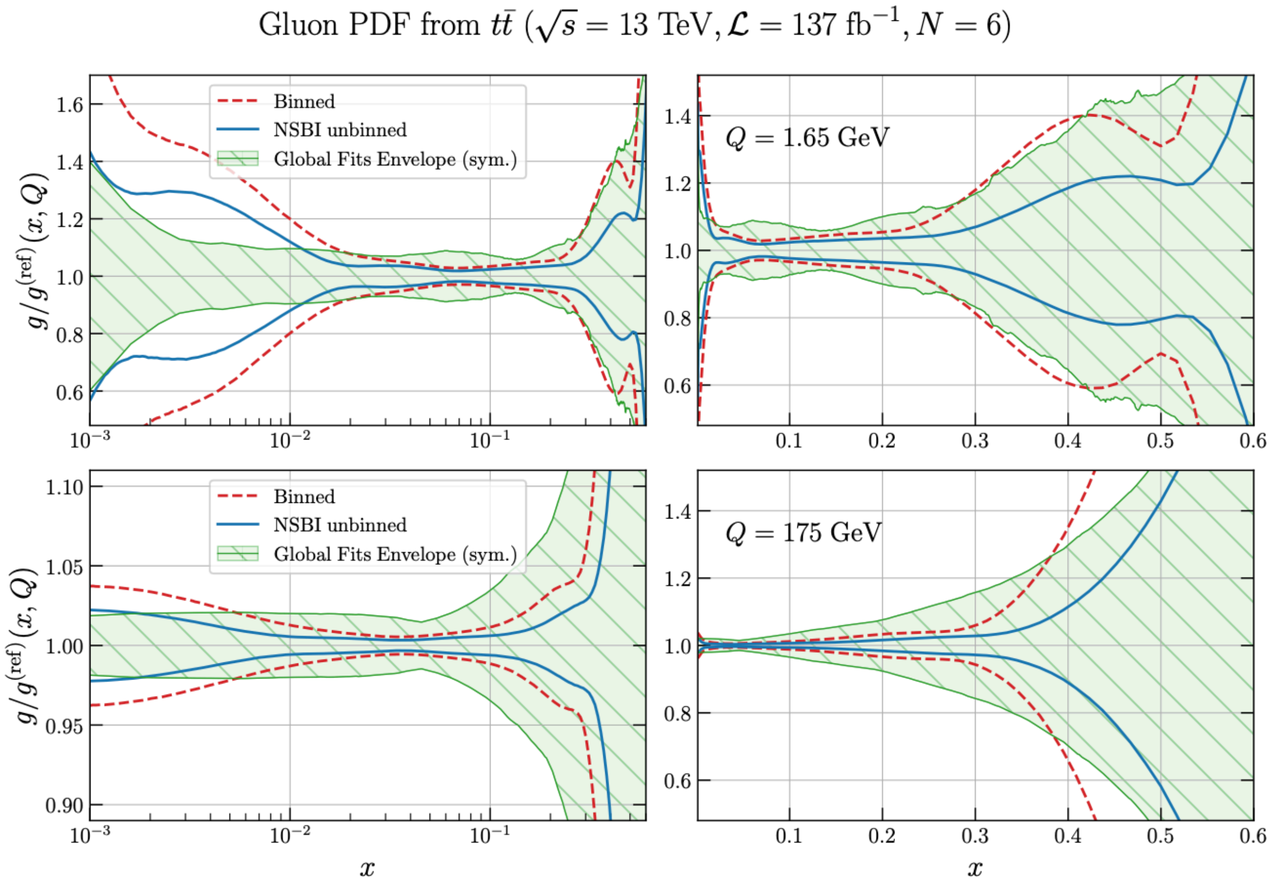


$$B^{(l)} = \frac{1}{N_{\rm data}} \sum_{i,j=1}^{N_{\rm data}} \left(\mathbb{E}_\epsilon T_i[f_k^{(l)}] - D_i^{L_0} \right) (C_{\rm PDF})_{ij}^{-1} \left(\mathbb{E}_\epsilon T_j[f_k^{(l)}] - D_j^{L_0} \right)$$

- Model-averaged gluon successfully recovers underlying law within estimated uncertainties.
- Posterior from average of models with best evidence faithfully reproduces statistical distribution of the data
- Under- and over-parametrised models systematically underestimate uncertainties in data region.

Gluon PDF from unbinned $t\bar{t}$ production at the LHC

Barrué et. al., ArXiv: 2604.13157



Linear PDF model with N = 6 parameters

Conclusions/Outlook

Summary / Conclusions

What we showed

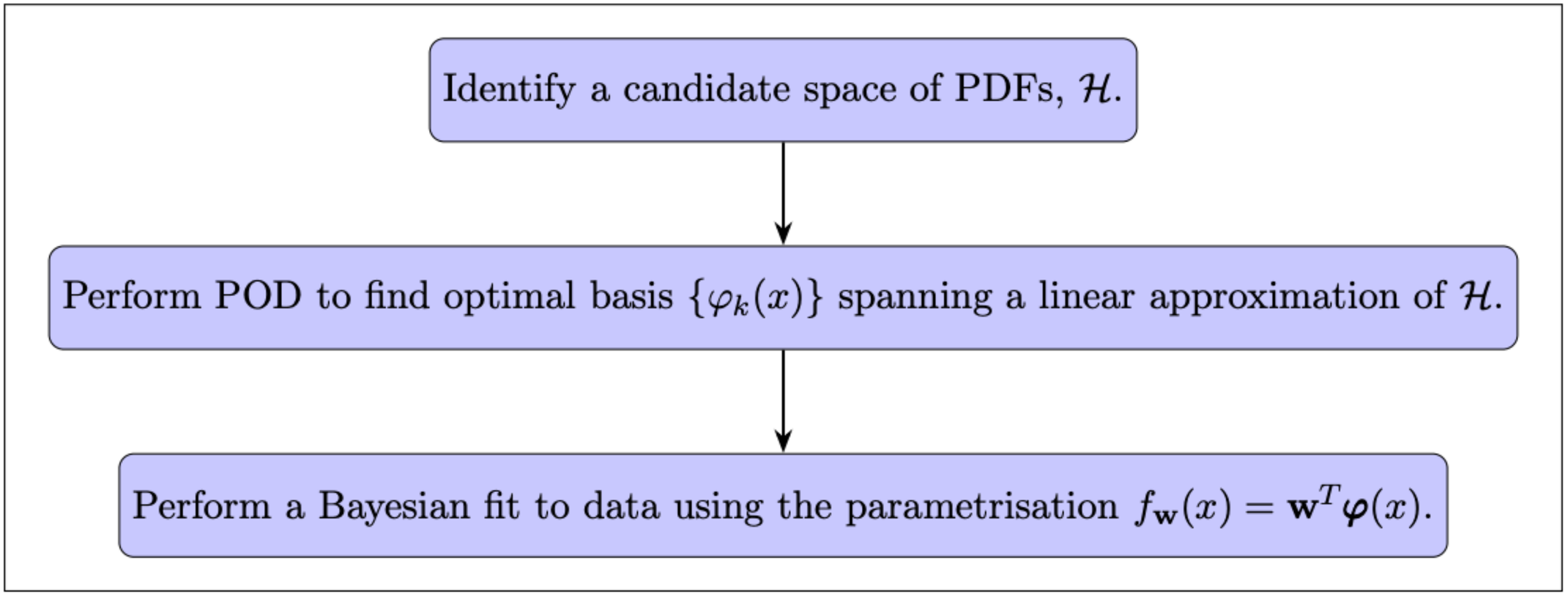
- Standard UQ methods (MC replica, Hessian) break down for **non-linear data** therefore a new approach is needed
- A linear PDF model built from POD of a large sample of NN initialisations
- Linear forward map + Gaussian data -> exact analytical posterior (no need for MCMC)
- Closure tests: confirm the Bayesian evidence identifies the right model complexity and uncertainty bands are statistically faithful
- The linear model enables neural SBI -> gluon determination from $t\bar{t}$ data

What comes next

- Full global analysis: DIS + Hadronic data
- Combined PDF + SM / BSM parameters extraction

Backup

A linear PDF parametrisation



Bayesian Workflow

Bayesian linear regression

Vast majority of the data used in PDF fits is “linear”

$$\mathbf{T} = \sum_{\text{partons}} \int dx\, f_q(x, Q^2)\hat{\sigma}(x, Q^2)$$

In this special case the forward map is also linear in the PDFs parameters!

$$\mathbf{T}(\mathbf{w}) = X\mathbf{w} + \epsilon$$

If data is Gaussian

$$\mathbf{D} \sim \mathcal{N}(\mathbf{D}_0, \Sigma)$$

Analytic posterior distribution

$$p(\mathbf{w}\,|\,\mathbf{D}_0) \sim \mathcal{N}(\hat{\mathbf{w}}, (X^T\Sigma^{-1}X)^{-1}), \qquad \hat{\mathbf{w}} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}\mathbf{D}_0$$

Bayesian linear regression

Given a model $\mathcal{M}_k$ the evidence is defined as

$$Z = p(\mathbf{D}_0 \,|\, \mathcal{M}_k) = \int d\mathbf{w}\; p(\mathbf{w} \,|\, \mathbf{D}_0, \mathcal{M}_k) p(\mathbf{w})$$

For a Gaussian posterior we can use the Laplace approximation

$$\ln Z = -\frac{1}{2}\chi^2 + \frac{N}{2}\ln(2\pi) + \ln\left(\frac{\sqrt{|(X^T\Sigma^{-1}X)^{-1}|}}{\prod_i (b_i - a_i)}\right)$$

Bayesian linear regression

Given a model $\mathcal{M}_k$ the evidence is defined as

$$Z = p(\mathbf{D}_0 \,|\, \mathcal{M}_k) = \int d\mathbf{w}\; p(\mathbf{w} \,|\, \mathbf{D}_0, \mathcal{M}_k) p(\mathbf{w})$$

For a Gaussian posterior we can use the Laplace approximation

$$\ln Z = -\frac{1}{2}\chi^2 + \frac{N}{2}\ln(2\pi) + \ln\left(\frac{\sqrt{|(X^T\Sigma^{-1}X)^{-1}|}}{\prod_i (b_i - a_i)}\right)$$

Favours models that fit well data

Bayesian linear regression

Given a model $\mathcal{M}_k$ the evidence is defined as

$$Z = p(\mathbf{D}_0 \,|\, \mathcal{M}_k) = \int d\mathbf{w}\; p(\mathbf{w} \,|\, \mathbf{D}_0, \mathcal{M}_k) p(\mathbf{w})$$

For a Gaussian posterior we can use the Laplace approximation

$$\ln Z = -\frac{1}{2}\chi^2 + \frac{N}{2}\ln(2\pi) + \ln\left(\frac{\sqrt{|(X^T\Sigma^{-1}X)^{-1}|}}{\prod_i (b_i - a_i)}\right)$$

Favours models that fit well data

penalises models with too many parameters

Non-linear regression

When the data is “non-linear”

$$T = \sum_{\text{partons}} \int dx_1 dx_2 f_{q1}(x_1, Q^2) f_{q2}(x_2, Q^2) \hat{\sigma}(x_1, x_2, Q^2)$$

We need MCMC to sample from the parameter space, eg Nested Sampling that also computes Z

In PDF fits we also impose other non-linear constraints: Positivity and Integrability

→ added in the form of penalties to the likelihood

Bayesian model average

Having fixed a POD basis we can explore multiple models $\mathscr{M}_k, k \in \{1,\dots,N\}$ with different number of basis elements

At the end we can average over all of them as

$$p(\mathbf{f}_{POD} \,|\, \mathbf{D}_0) = \sum_k p(\mathbf{D}_0 \,|\, \mathbf{f}_{POD}, \mathscr{M}_k) p(\mathscr{M}_k \,|\, \mathbf{D}_0)$$

And probability of the model given by

$$p(\mathscr{M}_k \,|\, \mathbf{D}_0) = \frac{p(\mathbf{D}_0 \,|\, \mathscr{M}_k)}{\sum_l p(\mathbf{D}_0 \,|\, \mathscr{M}_l)}$$

Choice of Prior

Uniform prior

$$f_{POD}(x, Q^2) = \xi_0(x, Q^2) + \sum_{i=1}^{N} w_i \left(\xi_i(x, Q^2) - \xi_0(x, Q^2)\right)$$

f_{POD} is linear in the w_i parameters → uniform prior in w_i results in uniform prior in f_{POD} !

However, in certain cases we have a much better choice

Choice of Prior

Bayesian Update

However, in certain cases we have a much better choice

$$\mathbf{D} \sim \mathcal{N}(\mathbf{T}(w), \Sigma), \text{ with } \Sigma = \Sigma_1 \oplus \Sigma_2, \mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2)$$

→ Fit on $\mathbf{d}_1$ yields a conditional distribution: $p(w \mid \mathbf{d}_1)$

$$p(w \mid \mathbf{d}_1) = \frac{\pi(w)\ \exp(-\frac{1}{2}||\mathbf{d}_1 - \mathbf{t}_1(w)||^2_{\Sigma_1})}{\int dw\ \pi(w)\ \exp(-\frac{1}{2}||\mathbf{d}_1 - \mathbf{t}_1(w)||^2_{\Sigma_1})}$$

→ Fit model to d_2 using $p(c \mid d_1)$ as prior

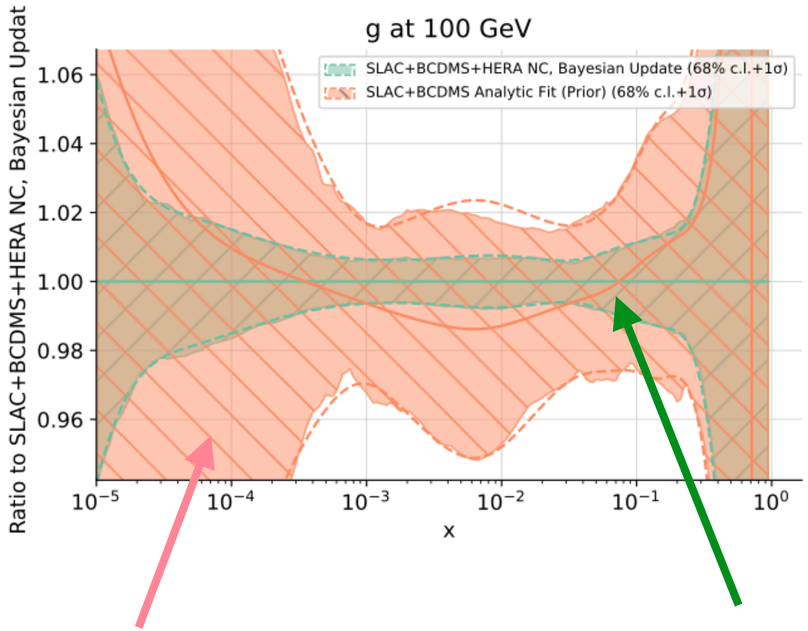
$$p(w \mid \mathbf{D}_0) = \frac{p_{\mathbf{d}_1}(w \mid \mathbf{d}_1)\ \exp(-\frac{1}{2}||\mathbf{d}_2 - \mathbf{t}_2(w)||^2_{\Sigma_2})}{\int dw\ p_{\mathbf{d}_1}(w \mid \mathbf{d}_1)\ \exp(-\frac{1}{2}||\mathbf{d}_2 - \mathbf{t}_2(w)||^2_{\Sigma_2})}$$

Choice of Prior

Bayesian Update

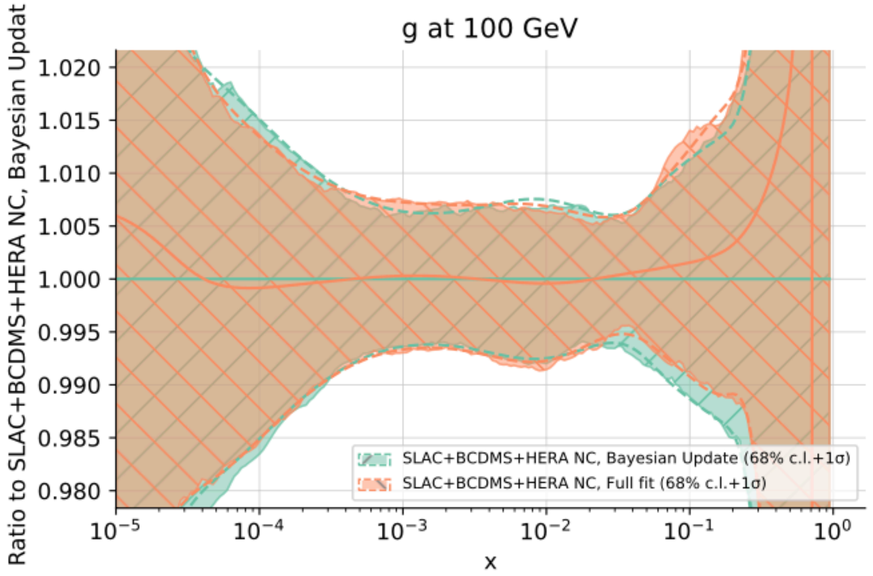
Example: consider data from (SLAC, BCDMS) + (HERA NC), (uncorrelated)

Fit SLAC + BCDMS first then use it as prior



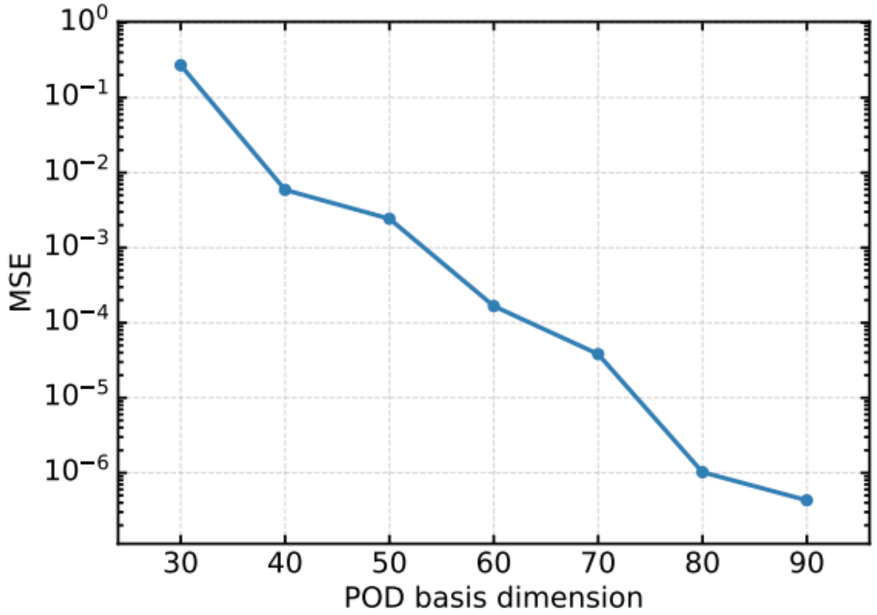
Prior distribution (SLAC+BCDMS fit)

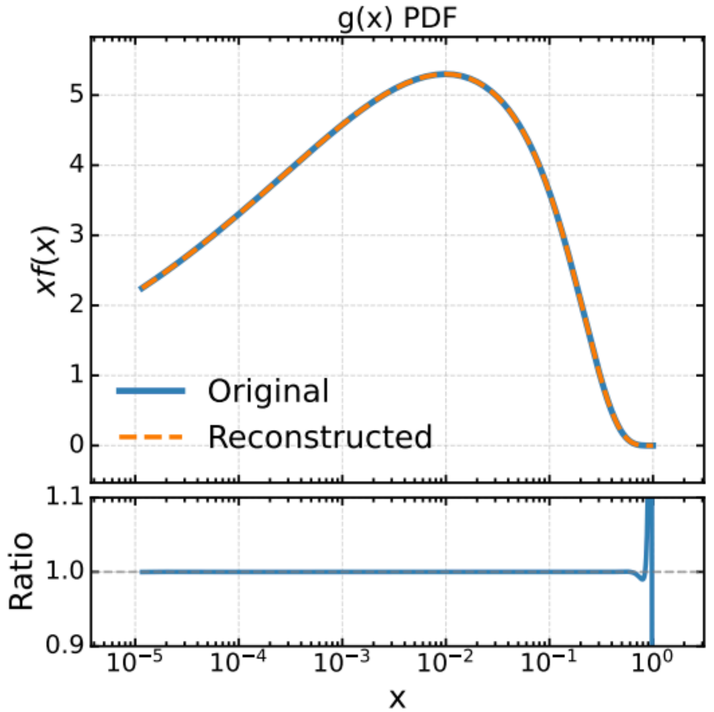
Full fit including HERA NC data



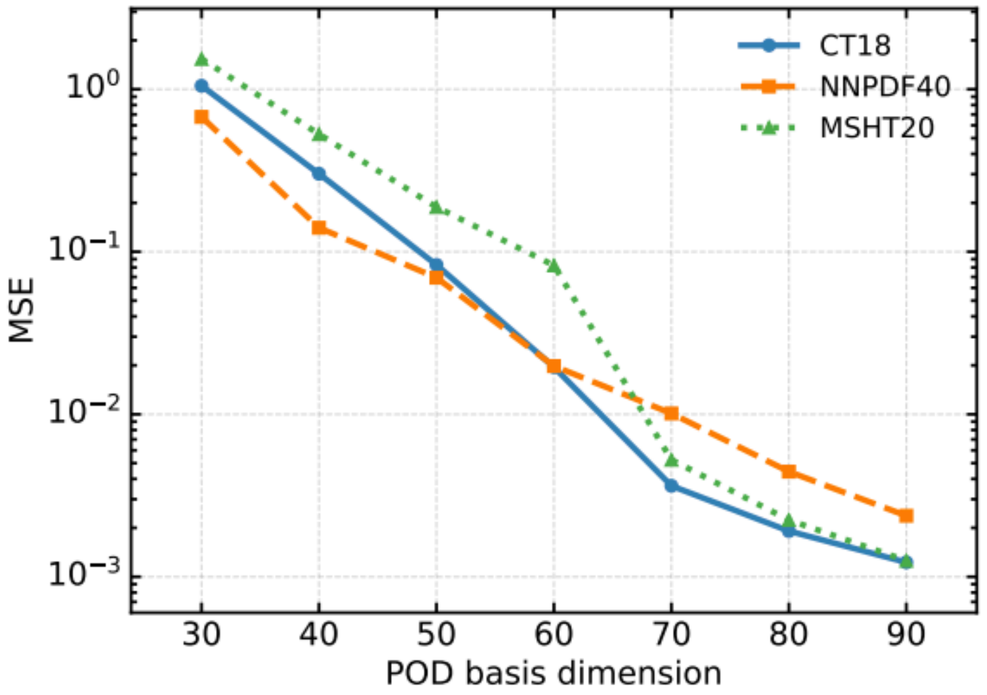
Comparison between bayesian update and uniform prior

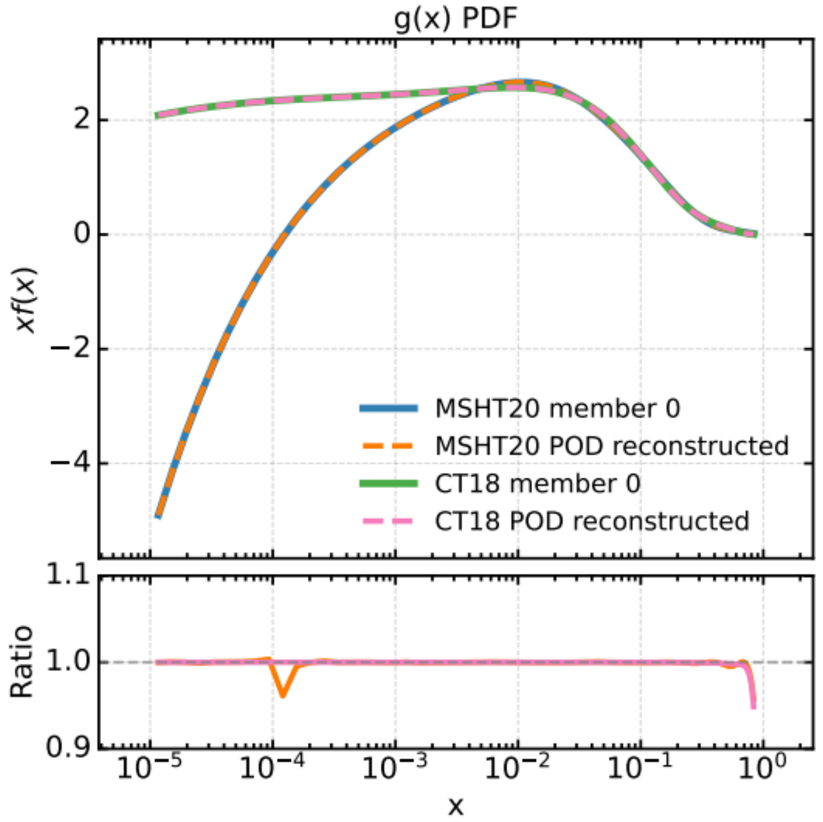
Completeness





Generalisation





Monte Carlo replica method

Suppose that experimental data is distributed according to a multivariate normal

$$\boldsymbol{d} \sim \mathcal{N}(\boldsymbol{t}(\boldsymbol{c}), \Sigma), \quad \Sigma \in \mathbb{R}^{N_{\text{dat}} \times N_{\text{dat}}}, \quad \boldsymbol{t} : \mathbb{R}^{N_{\text{param}}} \to \mathbb{R}^{N_{\text{dat}}}$$

Bayesian posterior

$$P_B(\boldsymbol{c} \,|\, \boldsymbol{d}_0) \propto \pi(\boldsymbol{c}) \ \exp\left(-\frac{1}{2}\chi^2(\boldsymbol{t}(\boldsymbol{c}), \boldsymbol{d}_0)\right)$$

Monte Carlo replica method:

- Pseudodata distribution $\boldsymbol{d}_p \sim \mathcal{N}(\boldsymbol{d}_0, \Sigma)$
- “Best fit parameter"
  $\boldsymbol{c}_p(\boldsymbol{d}_p) = \text{argmin}_{\boldsymbol{c}}(\boldsymbol{d}_p - \boldsymbol{t}(\boldsymbol{c}))^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{t}(\boldsymbol{c}))$

How is $\boldsymbol{c}_p(\boldsymbol{d}_p)$ distributed?

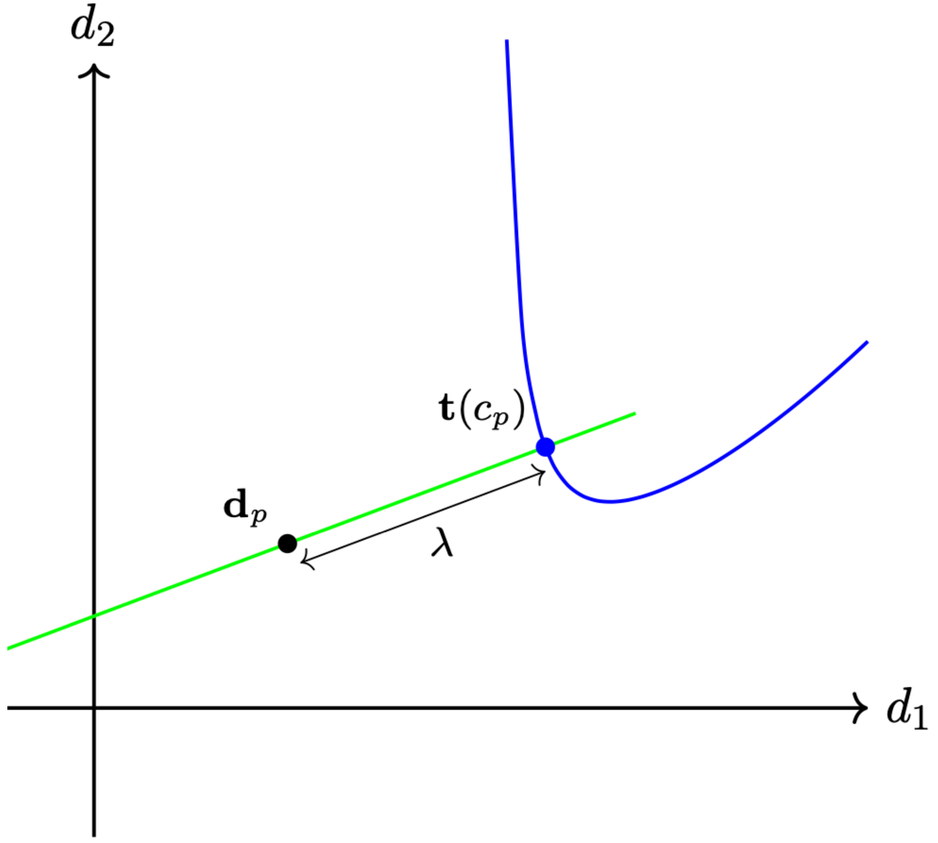
Monte Carlo Posterior

$\boldsymbol{c}_p(\boldsymbol{d}_p)$ Is a random function of $\boldsymbol{d}_p$

$$P_{\mathrm{MC}}(\boldsymbol{c}) \propto \int d^{N_{\mathrm{dat}}}\boldsymbol{d}_p \;\; \delta\left(\boldsymbol{c} - \boldsymbol{c}_p(\boldsymbol{d}_p)\right) \exp\left(-\frac{1}{2}(\boldsymbol{d}_p - \boldsymbol{d}_0)^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{d}_0)\right)$$

A general form for $P_{\mathrm{MC}}(\boldsymbol{c})$ can be found by performing a coordinate transformation

$$\boldsymbol{d}_p \rightarrow (\boldsymbol{c}_p, \boldsymbol{\lambda})$$



Monte Carlo Posterior

Linear theory

See also L. D. Debbio, T. Giani, M. Wilson [2111.05787]

$$\boldsymbol{t}(\boldsymbol{c}) = \boldsymbol{t}_0 + \boldsymbol{t}_{\text{lin}}\boldsymbol{c}$$

The Monte Carlo posterior

$$P_{\text{MC}}(\boldsymbol{c}) \propto \int d^{N_{\text{dat}}}\boldsymbol{d}_p \delta\left(\boldsymbol{c} - \boldsymbol{c}_p(\boldsymbol{d}_p)\right) \exp\left(-\frac{1}{2}(\boldsymbol{d}_p - \boldsymbol{d}_0)^T \Sigma^{-1} (\boldsymbol{d}_p - \boldsymbol{d}_0)\right)$$

reduces to

$$P_{\text{MC}}(\boldsymbol{c}) \propto \exp\left(-\frac{1}{2}\chi^2_{\mathbf{d}_0}(\mathbf{c})\right)$$

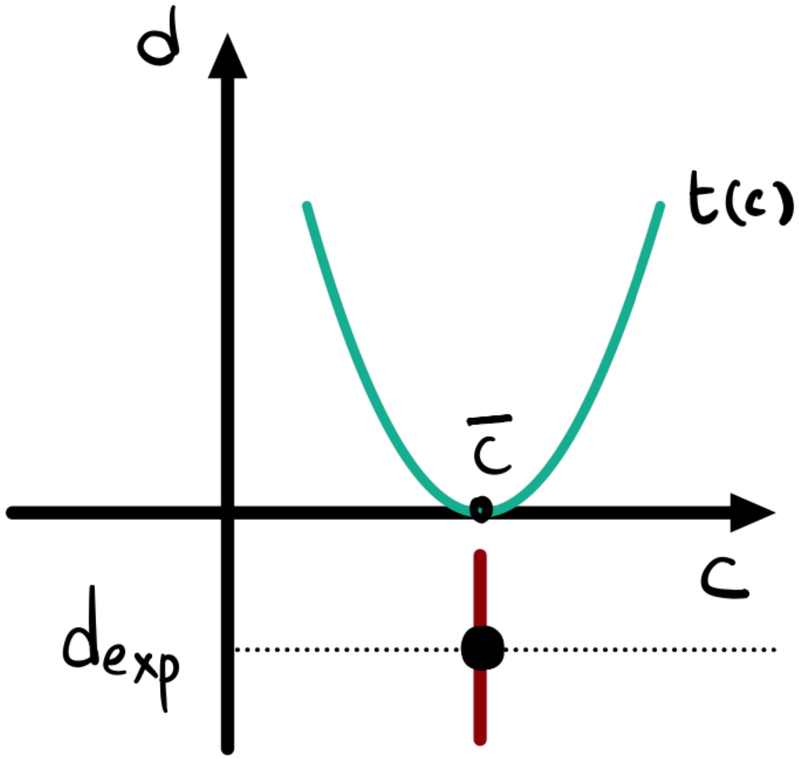
For a linear theory, the Bayesian and Monte Carlo approaches coincide (wide flat prior)

Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2,\ t_{\text{quad}} > 0$$

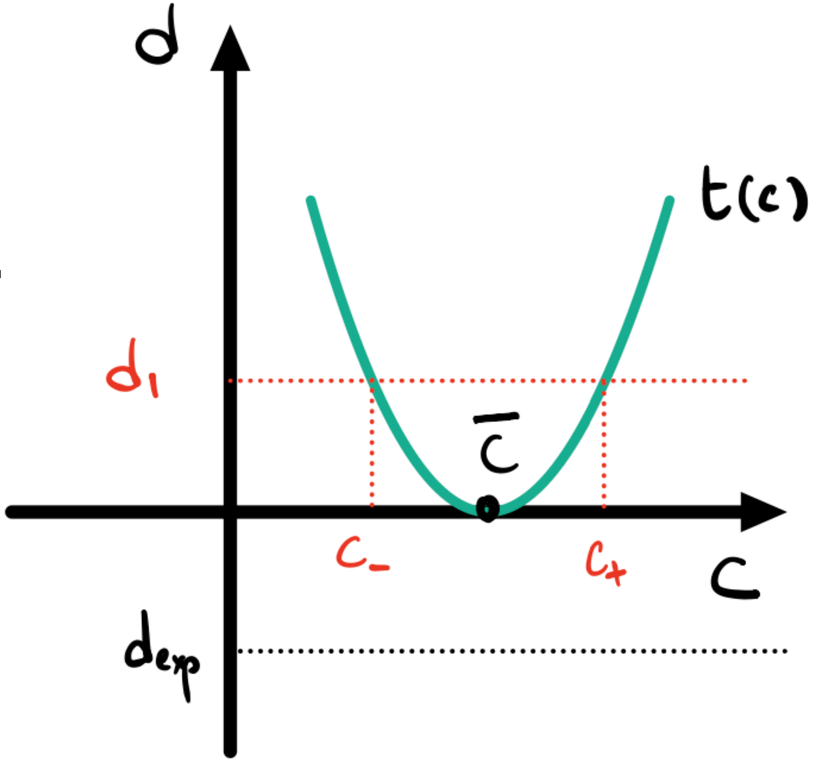


Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2,\ t_{\text{quad}} > 0$$



$$\chi^2 = \frac{(t(c) - d_{\text{exp}})^2}{\sigma^2_{\text{exp}}},\ \ \Delta\chi^2 = \chi^2 - \chi^2_{\text{min}} = 1 \quad \rightarrow [c_-, c_+]$$

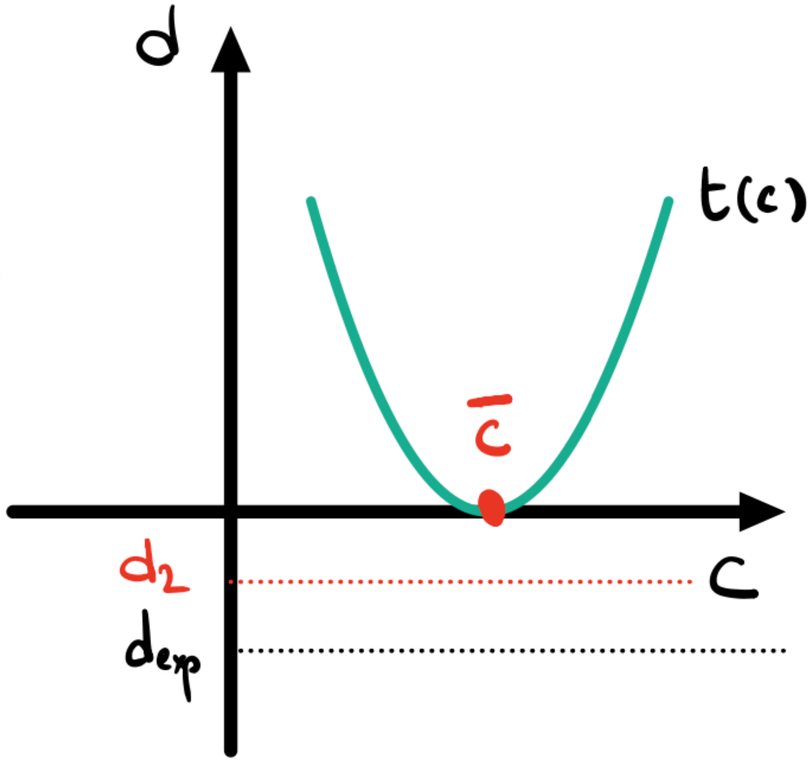
Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}}c + t_{\text{quad}}c^2,\ t_{\text{quad}} > 0$$

$$c = -\frac{t_{\text{lin}}}{2t_{\text{quad}}}$$

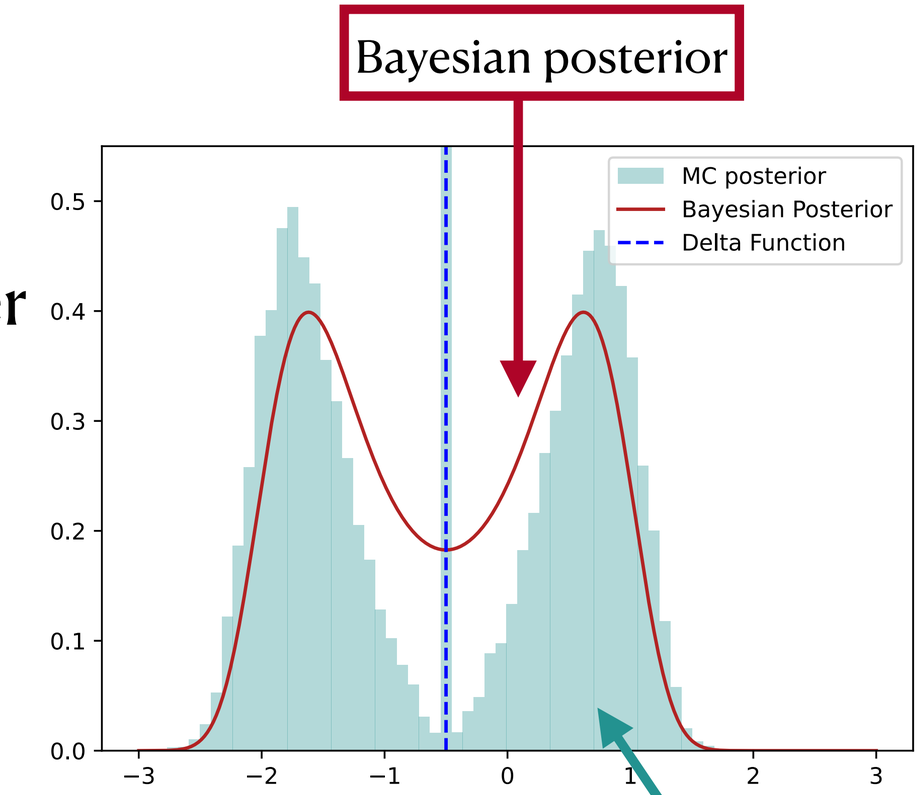


Monte Carlo Posterior

Quadratic Model

Let's consider one datapoint, one theory parameter

$$t(c) = t_0 + t_{\text{lin}} c + t_{\text{quad}} c^2,\ t_{\text{quad}} > 0$$



$$P_{\text{MC}}(c) \propto \delta\left(c + \frac{t_{\text{lin}}}{2t_{\text{tquad}}}\right) \int_{-\infty}^{t_{\text{min}}} dd_p \exp\left(-\frac{1}{2}\frac{(d_0 - d_p)^2}{\sigma^2}\right) + 2\,|\,t_{\text{lin}} + 2ct_{\text{quad}}\,|\exp\left(-\frac{1}{2}\frac{(d_0 - t(c))^2}{\sigma^2}\right)$$

Independent on d_p: $\frac{\partial t(c)}{\partial c}$ has not full rank

2 roots

"Determinant" factor

Nested Sampling

General Idea

- Monte Carlo algorithm for computing an integral over a model parameter space
- Nested Sampling provides both the posterior samples as well as the marginalised likelihood Z

Bayes Rule

$$P(\Theta \mid D) = \frac{L(D \mid \Theta)\pi(\Theta)}{Z}$$

Marginalised Likelihood

$$Z = \int L(D \mid \Theta)\pi(\Theta)d\Theta$$

Nested Sampling

Algorithm

1. Initialisation: sample randomly from the prior N live points and compute the Likelihood at each point
2. Shrinkage: remove point with the lowest likelihood L_1
3. Likelihood Restricted Prior Sampling: sample new point from prior with Likelihood $> L_1$
4. Iterate

Iteration i reduces integration volume by a factor $\delta V_i \approx \left(1 - \frac{1}{N}\right)^i \frac{1}{N}$,

The integral Z is simply $Z \approx \sum_i \delta V_i \times L_i$

Termination: when $\delta V_i \times L_i$ contributions to Z are negligible

Nested Sampling

Summary

1. It explores the parameter space globally;
2. it handles multi-modal distributions well;
3. it initialises and terminates at a well defined point -> no supervision;
4. it provides both marginal likelihood and posterior samples, hence allowing for Bayesian model selection