Uncertainty Quantification: discussion

CNF GPD workshop

JLab

06/13/23

Instituto de Física
UNAM

CONACYT
Consejo Nacional de Ciencia y Tecnología

dgapa
Dirección General de Asuntos
del Personal Académico

Aurore Courtoy

Instituto de Física
Universidad Nacional Autónoma de México (UNAM)

Uncertainty quantification for GPDs

Global analysis of GPD differs from that of PDFs, but can nonetheless be inspired by some aspects at the forefront of the phenomenological extraction of functions of one (or more) variables.

GPD extraction is the definition of “a very complex inverse problem.”

⇨ strong points: (analytical) theoretical constraints to exploit — *GPD representations*.

⇨ less strong points: not quite at the “accuracy vs. precision” level, subtleties from integrant and interplay with evolution — *what are shadow GPDs in UQ language?*

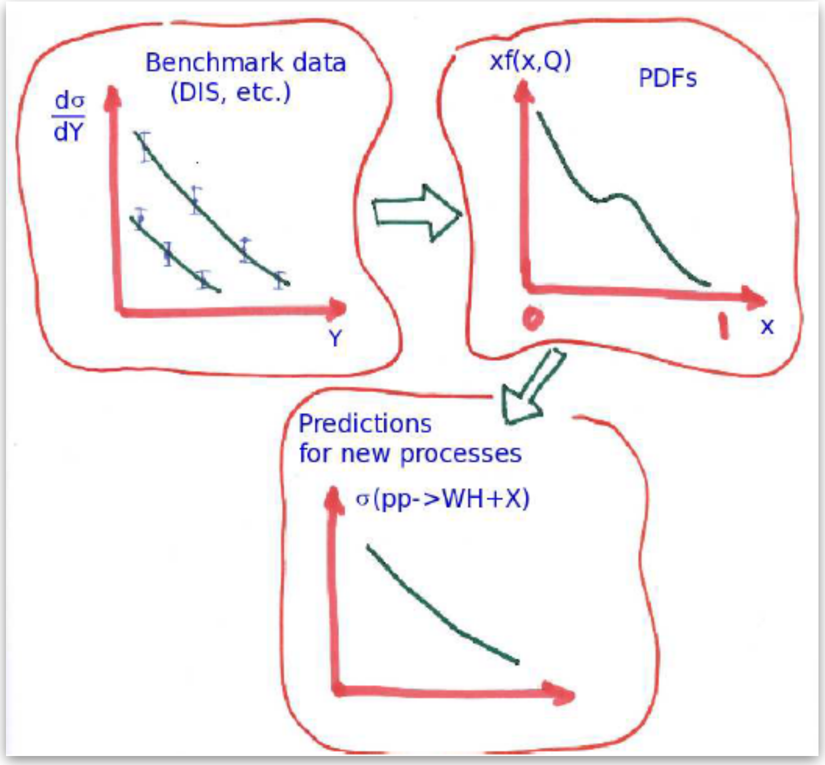
⇨ points that were put aside: dependence on ξ is all that’s left in the CFF —*DGLAP/ERBL differences.*

Constraints

are usually fulfilled through **parametrization, Lagrange multipliers and/or prior conditions**.
To some extent, all constraints are biases, and need to be treated carefully.

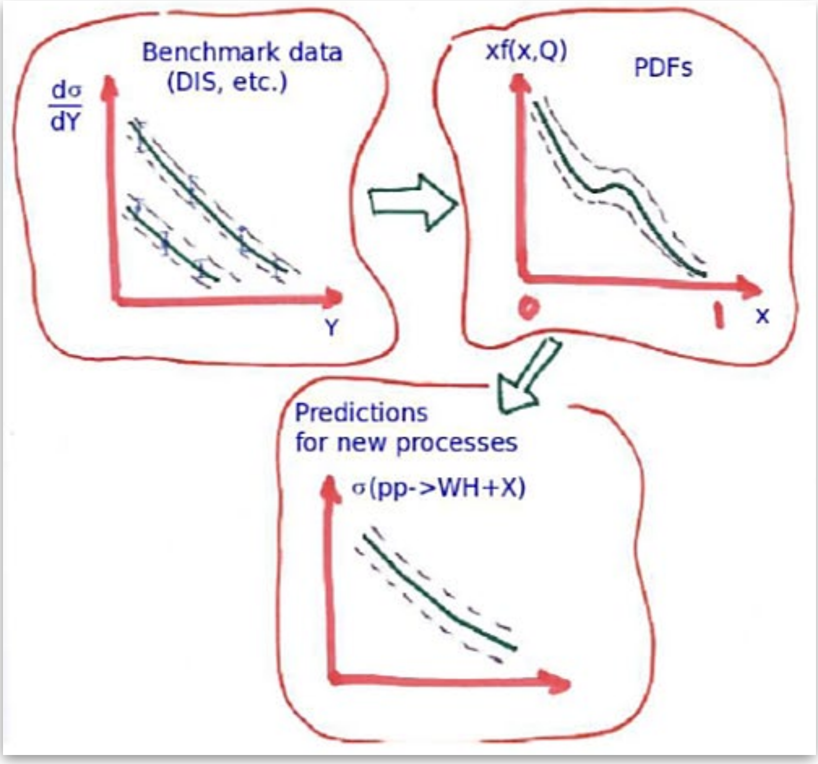
The shape of parton distributions

Parton distributions are functions of the momentum fraction x, they are extracted from data that are sensitive to specific PDF flavors, etc.



⇨ finding the shape in x is the goal of PDF analyses

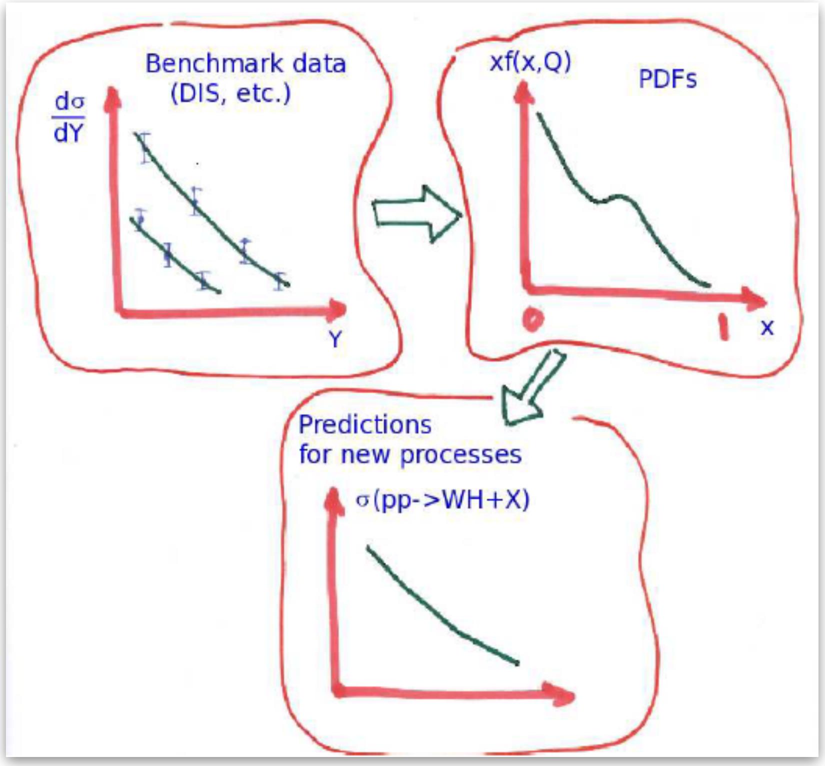
Uncertainty propagates from data and methodology to the PDF determination.
There are two classes of them,



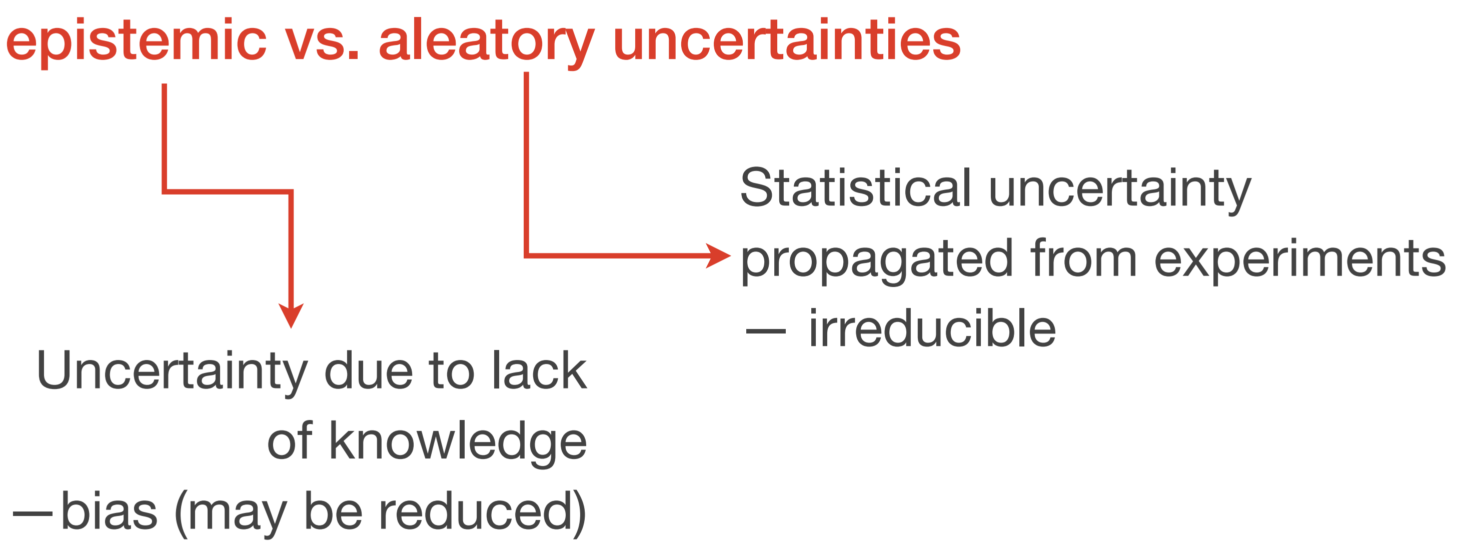
The shape of parton distributions

Parton distributions are functions of the momentum fraction x, they are extracted from data that are sensitive to specific PDF flavors, etc.

➩ finding the shape in x is the goal of PDF analyses



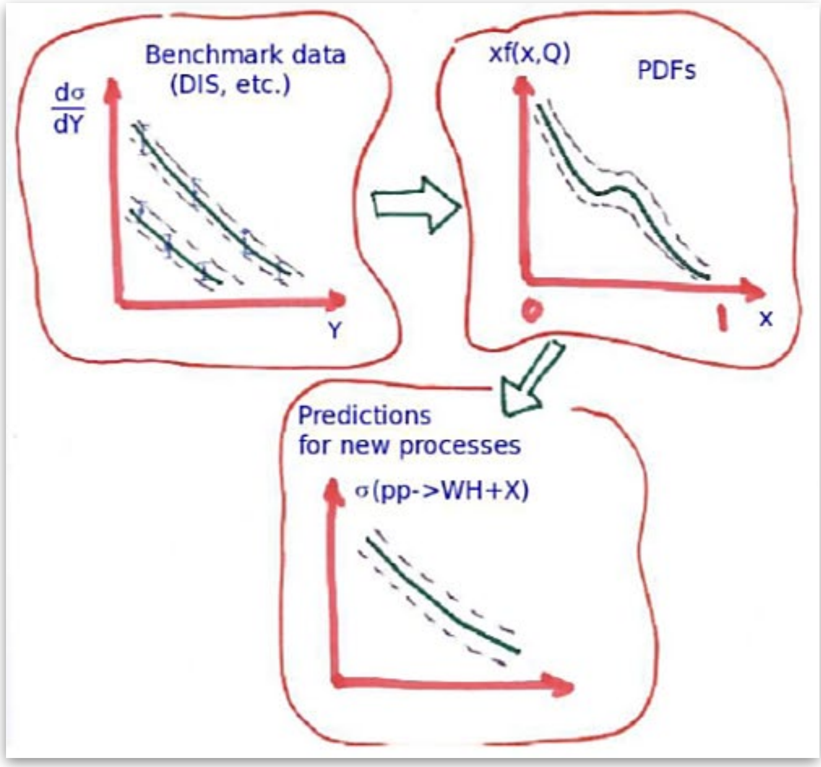
Uncertainty propagates from data and methodology to the PDF determination.
There are two classes of them,



epistemic vs. aleatory uncertainties

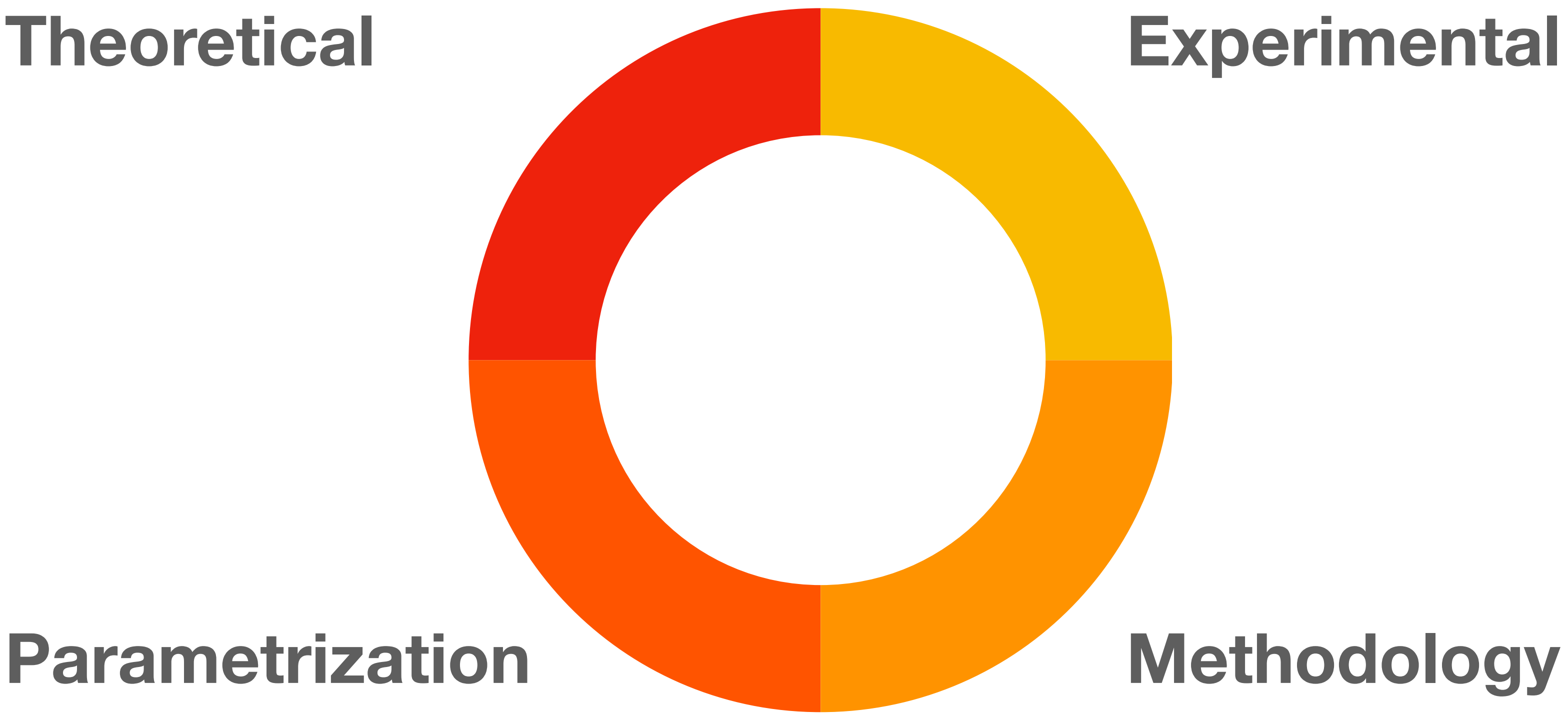
Statistical uncertainty propagated from experiments — irreducible

Uncertainty due to lack of knowledge —bias (may be reduced)



On uncertainty quantification

During today's sessions, "model" seemed to include all 4 but the experimental one!



Theoretical

Experimental

Parametrization

Methodology

In all four categories of uncertainties, we can further distinguish *PDF fitting accuracy* from *PDF sampling accuracy*.

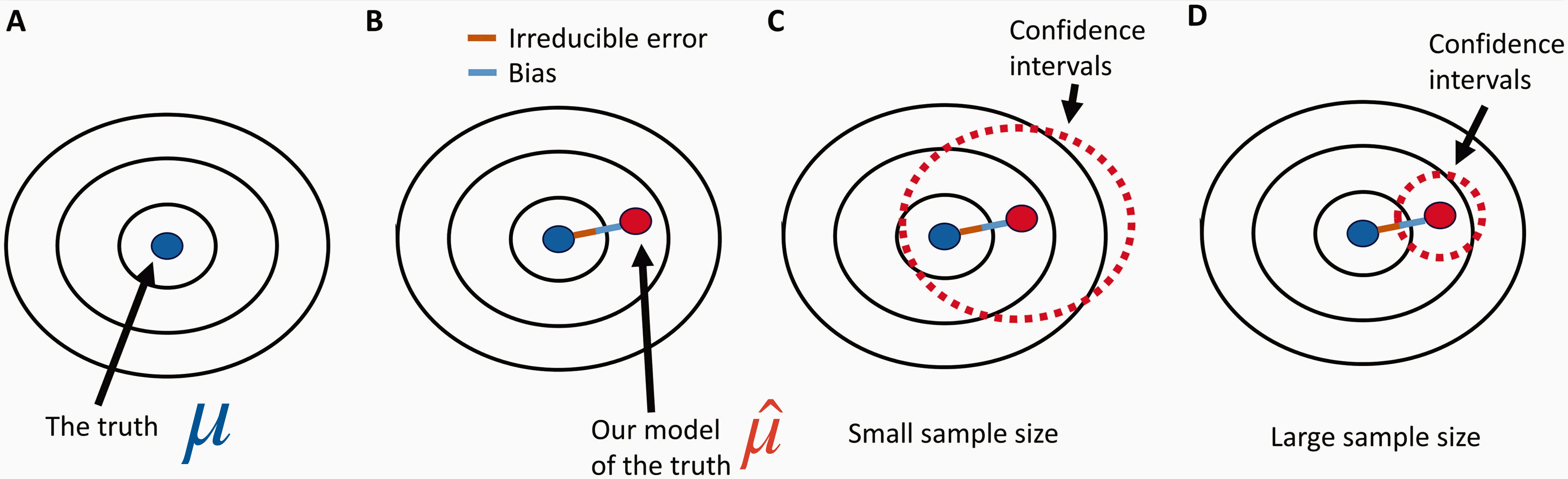
Goodness-of-fit applies to an individual best fit.

[Kovarik et al, Rev.Mod.Phys. 92 (2020)]

Sampling accuracy applies to the PDF/… ensemble.

[AC et al, Phys.Rev.D107 (2023)]

Sampling bias and big-data paradox



Pavlos Msaouel (2022)
Cancer Investigation, 40:7, 567-576

With an increasing size of sample $n \to \infty$, under a set of hypotheses, it is usually expected that the *deviation* on an observable decreases like $\left(\sqrt{n}\right)^{-1}$.

That's the law of large numbers.

What uncertainties keep us from including *the truth*, μ?

The law of large numbers disregards the *quality of the sampling*, 🟧 Irreducible error 🟦 Bias .

Xiao-Li Meng
The Annals of Applied Statistics
Vol. 12 (2018), p. 685

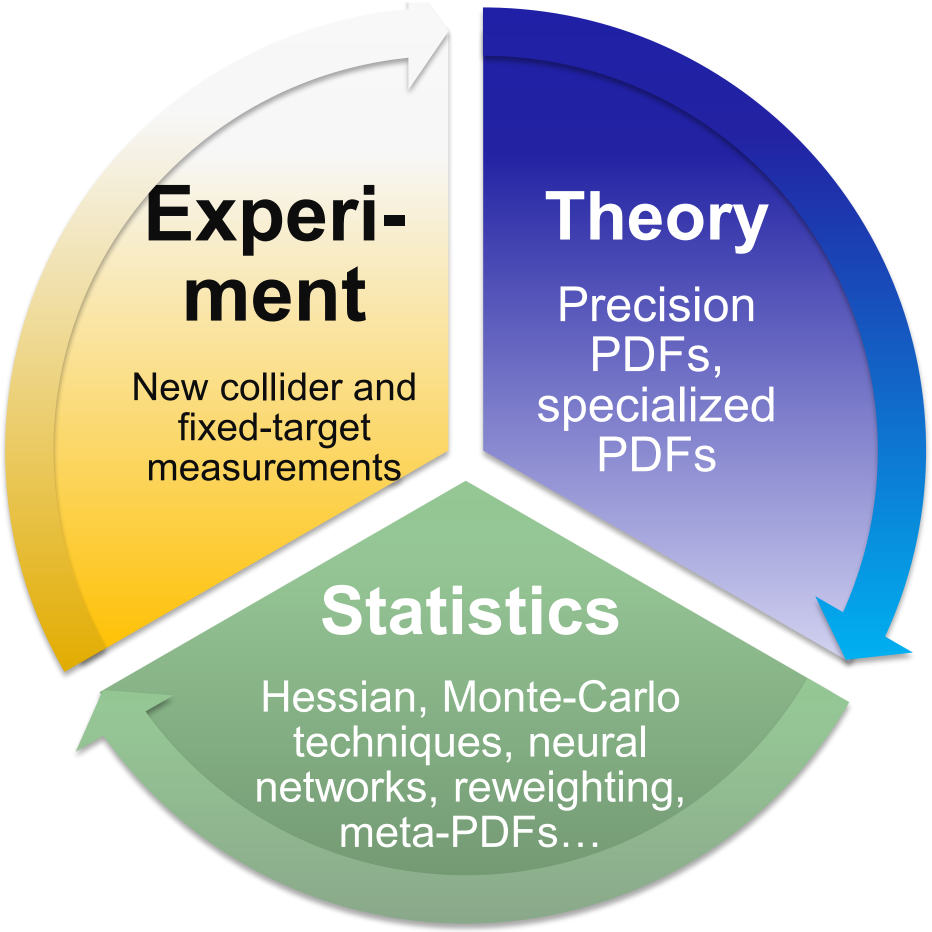
Sampling bias in PDF global analyses

There is a “data+sampling defect=confounding correlation” factor in global analyses.

Methodological choices are reflected in the epistemic uncertainty, including biases from sampling.

Priors, including choice of functional form or Bayesian *priors*, influence the sampling algorithm.

Representative sampling accounts for the confounding correlation, and can ultimately be used to optimize its contribution, e.g. through the study of largest effective dimensions.


Experi-ment
New collider and fixed-target measurements
Theory
Precision PDFs, specialized PDFs
Statistics
Hessian, Monte-Carlo techniques, neural networks, reweighting, meta-PDFs…

⇨ dimensionality reduction (effective dimensions) vs. phase space reduction (priors)

Hypothesis testing and parton distributions

Representative sampling

Acceptable functions

Tests of PDFs

Analytic conditions and acceptable bias to be adapted to GPDs.

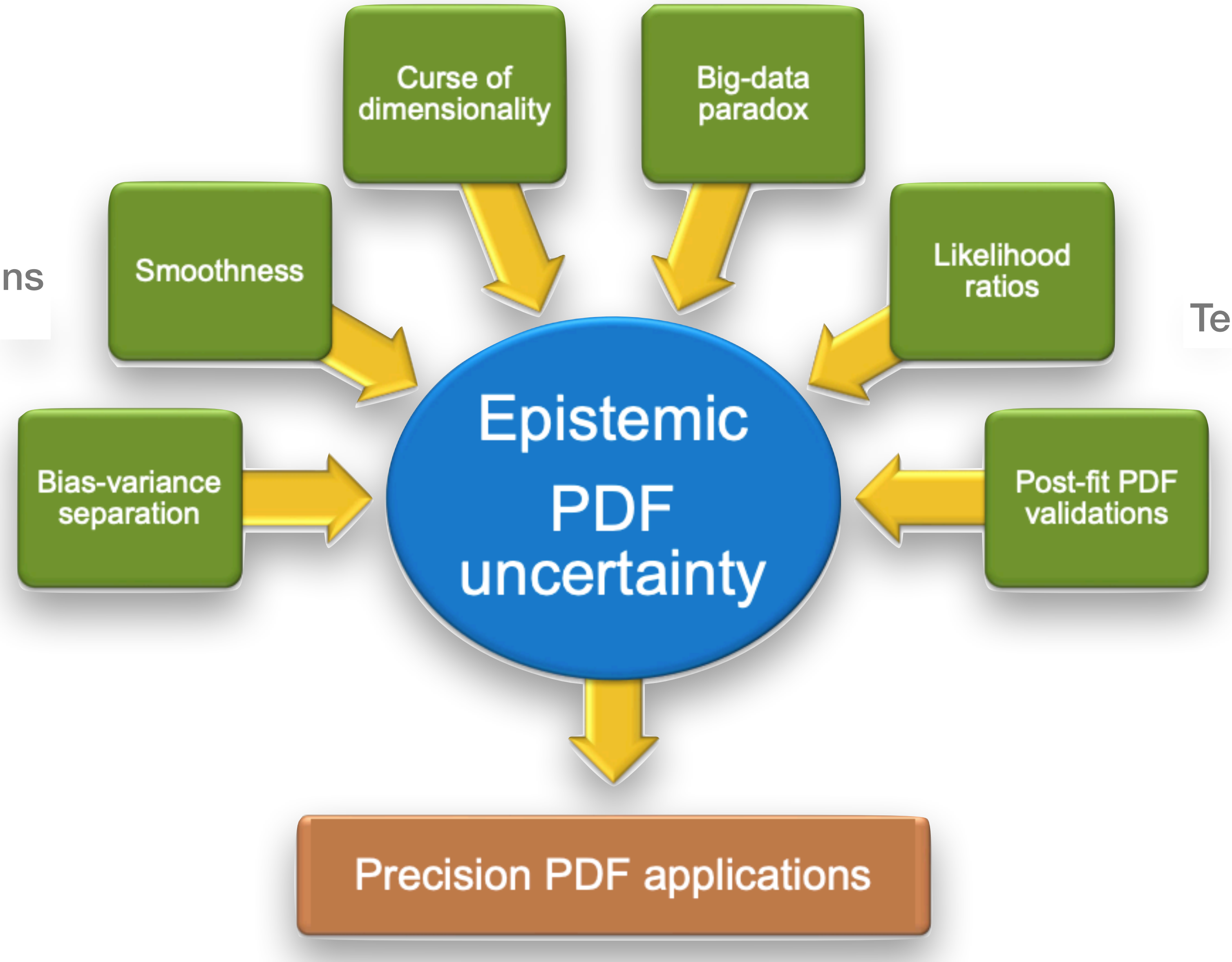


diagram by P. Nadolsky [DIS2023]

Likelihood and sampling — I

What is the adequate objective function for PDF analyses?

$$P(a|D) \propto P(D|a)\, P(a)$$

$$\Leftrightarrow \exp(-\chi^2_{\rm aug}/2) \propto \exp(-\chi^2/2)\, \exp(-\chi^2_{\rm prior}/2)$$

$$\Rightarrow \chi^2_{\rm aug} = \chi^2 + \chi^2_{\rm prior}$$

[Lepage et al., NPB Proc.Suppl.106(2002) 12-20]

Parameters: parameters of interest a and nuisance parameters.

Likelihood: “augmented” likelihood contains constraints/priors/penalties as well as the minimal likelihood. Identify priors on a.

⇨ to some extent, similar to Lagrange multipliers

Likelihood and sampling — II

On which basis are DFs accepted or rejected?

Likelihood ratios:
two replicas can be ordered according to their relative likelihood or relative prior.

$$\frac{P(T_2|D)}{P(T_1|D)} = \frac{P(D|T_2)}{P(D|T_1)} \times \frac{P(T_2)}{P(T_1)}$$

$\equiv r_{\text{posterior}}$ $\equiv r_{\text{likelihood}}$ $\equiv r_{\text{prior}}$

aleatory epistemic + aleatory probabilities

Prior: replica can be discarded based on $P(T_2) < P(T_1)$ even for $r_{likelihood} \sim 1$

Likelihood: replica can be accepted based on $r_{likelihood} = \dfrac{P(D\,|\,T_2)}{P(D\,|\,T_1)} \sim 1$ when $P(T_2) \sim P(T_1)$

Key role played by priors

Priors have been identified to reduce the phase space

Constraints fit exploits the benefits of well-controlled priors

“Constrained curve fitting,” [Lepage et al., Nucl.Phys.B Proc.Suppl.106(2002) 12-20] — lattice oriented

⇨ similarly for polarized PDF analysis [Benel et al, EPJC]

Solutions may be prejudiced by strong priors

Some publications show how strong priors have affected results (that has led to important claims)

⇨ Proton structure: “Parton distributions need representative sampling”+ communication with NNPDF

[CT, PRD107, 2023]

⇨ Neutrino physics: “Neutrino mass and mass ordering: no conclusive evidence for normal ordering”

[Stefano Gariazzo et al JCAP10(2022)010]

Uncertainty quantification for GPDs

Global analysis of GPD differs from that of PDFs, but can nonetheless be inspired by some aspects at the forefront of the phenomenological extraction of functions of one (or more) variables.

GPD extraction is the definition of “a very complex inverse problem.”

⇨ strong points: (analytical) theoretical constraints to exploit — *GPD representations*.

Can we exploit the “old-style” parametric generation of possible solutions (double dstr., dual param.)?

⇨ less strong points: not quite at the “accuracy vs. precision” level, subtleties from integrant and interplay with evolution — *what are shadow GPDs in UQ language?*

Can we think in terms of the bias-variance dilemma?

⇨ points that were put aside: dependence on ξ is all that’s left in the CFF—*DGLAP/ERBL differences.*

Can we learn from theory development for the pion (DA vs. PDF)?

Benchmark for UQ for GPDs

GPD sampling probably at least inclusive of the same items are forPDF sampling, that takes place over

experimental data sets,
parametrization forms,
hyperparameters,
settings of fits,
model approximations.
+ choice of likelihood (and treatment of syst. uncertainties).

Out-of-fit tests: are closure tests inclusive of sampling bias on the “model”?

⇨ Epistemic uncertainty can only be optimized if it is understood —though irreducible in certain cases.

Back up

A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

Step 1

The NNPDF4.0 Hessian set ($n = 50$) defines a coordinate system on a manifold corresponding to the largest variations of the PDF uncertainty —red dots and curve.

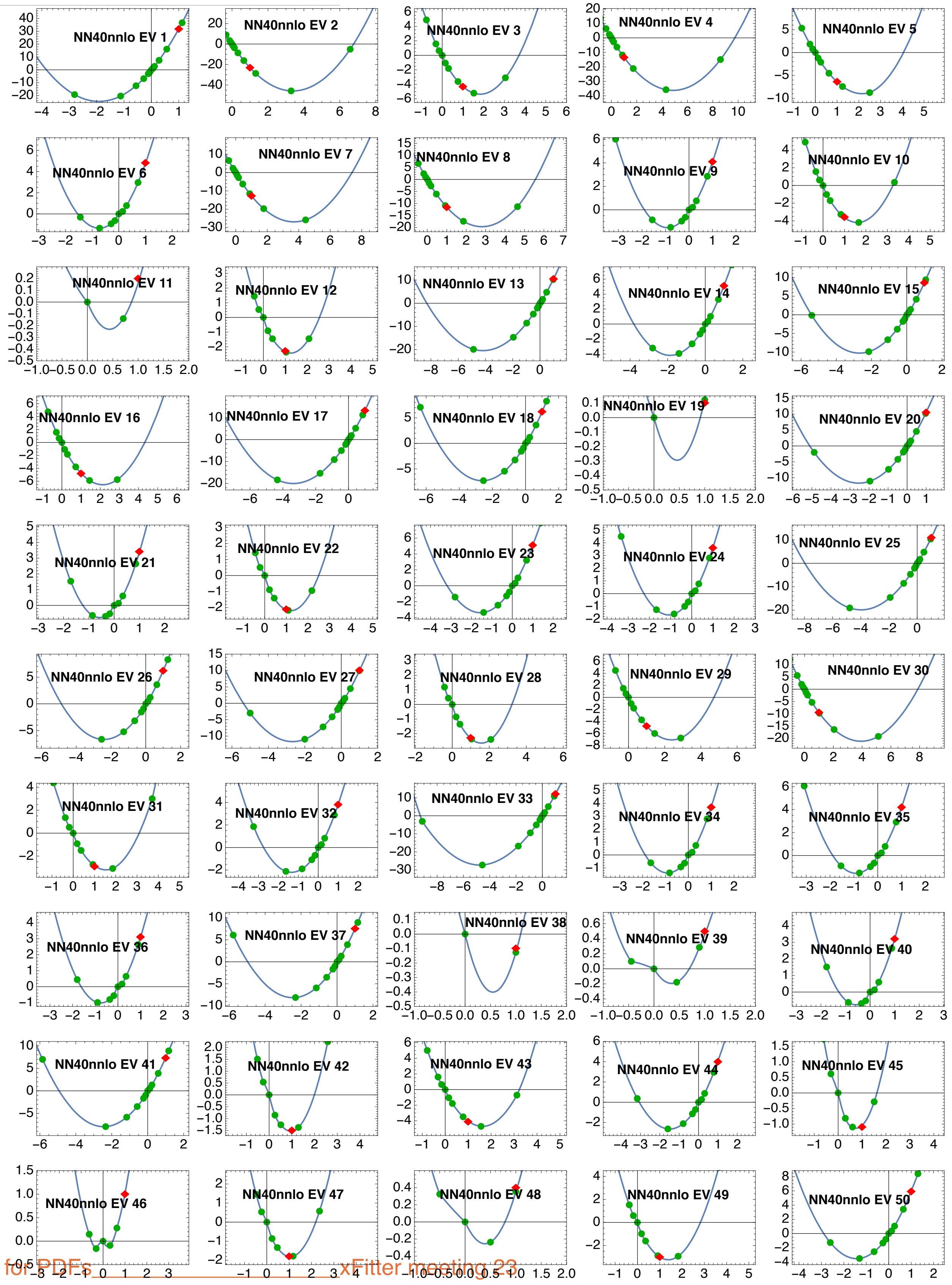
[NNPDF, 2109.02653]

Step 2

Using the public NNPDF code, scan χ^2_{tot} along the 50 EV directions to identify a hypercube corresponding to $\Delta\chi^2 \leq T^2$ (where $T^2 > 0$ is a user-selected value).

Lagrange multiplier scan confirms the approximate Gaussian profiles, but suggest that there exist solutions with lower χ^2 — green dots and blue curve.

No fitting involved.



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

Step 1

The NNPDF4.0 Hessian set ($n = 50$) defines a coordinate system on a manifold corresponding to the largest variations of the PDF uncertainty —red dots and curve.

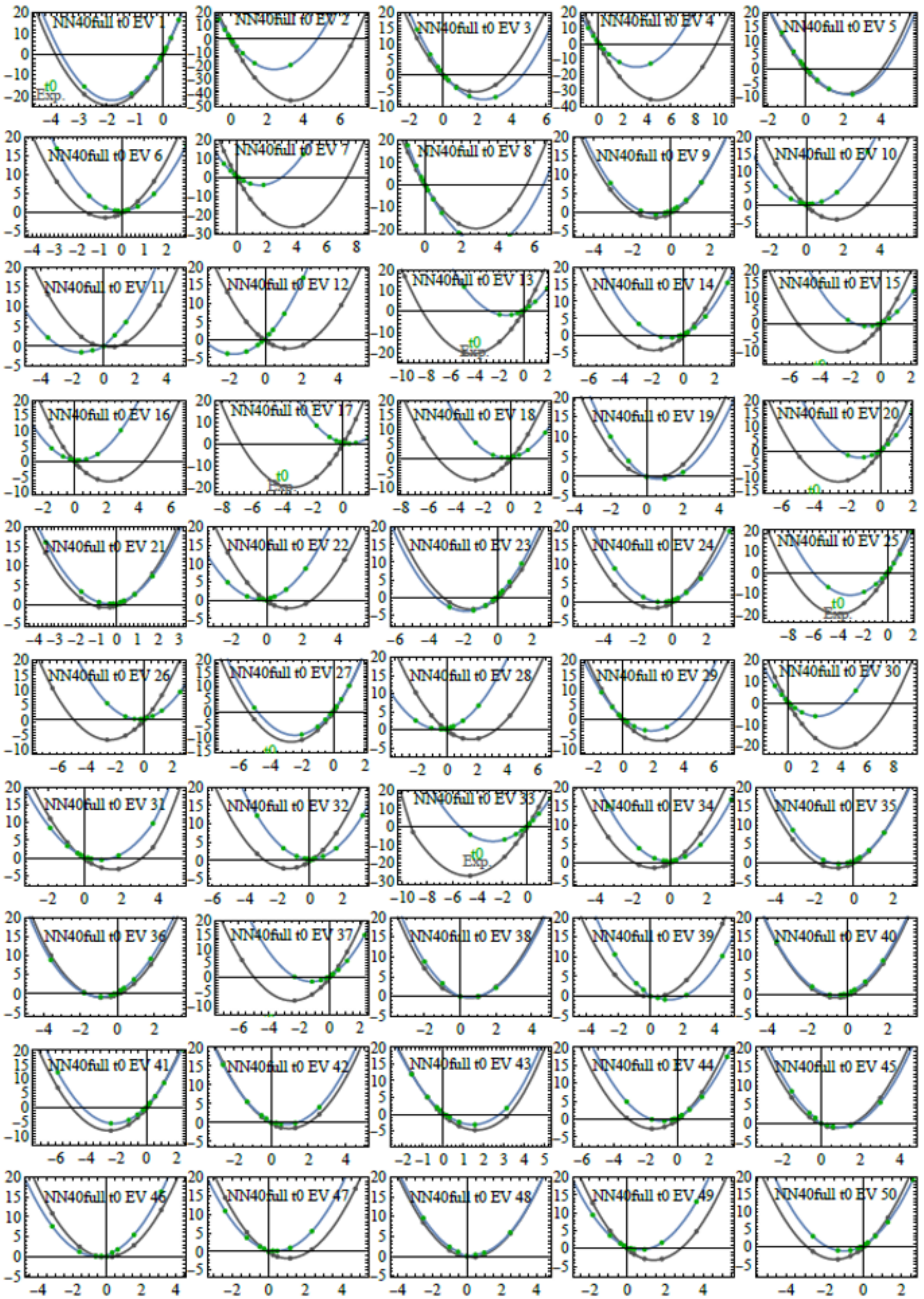
[NNPDF, 2109.02653]

Step 2

Using the public NNPDF code, scan χ^2_{tot} along the 50 EV directions to identify a hypercube corresponding to $\Delta\chi^2 \leq T^2$ (where $T^2 > 0$ is a user-selected value).

Lagrange multiplier scan confirms the approximate Gaussian profiles, but suggest that there exist solutions with lower χ^2 — green dots and blue curve.

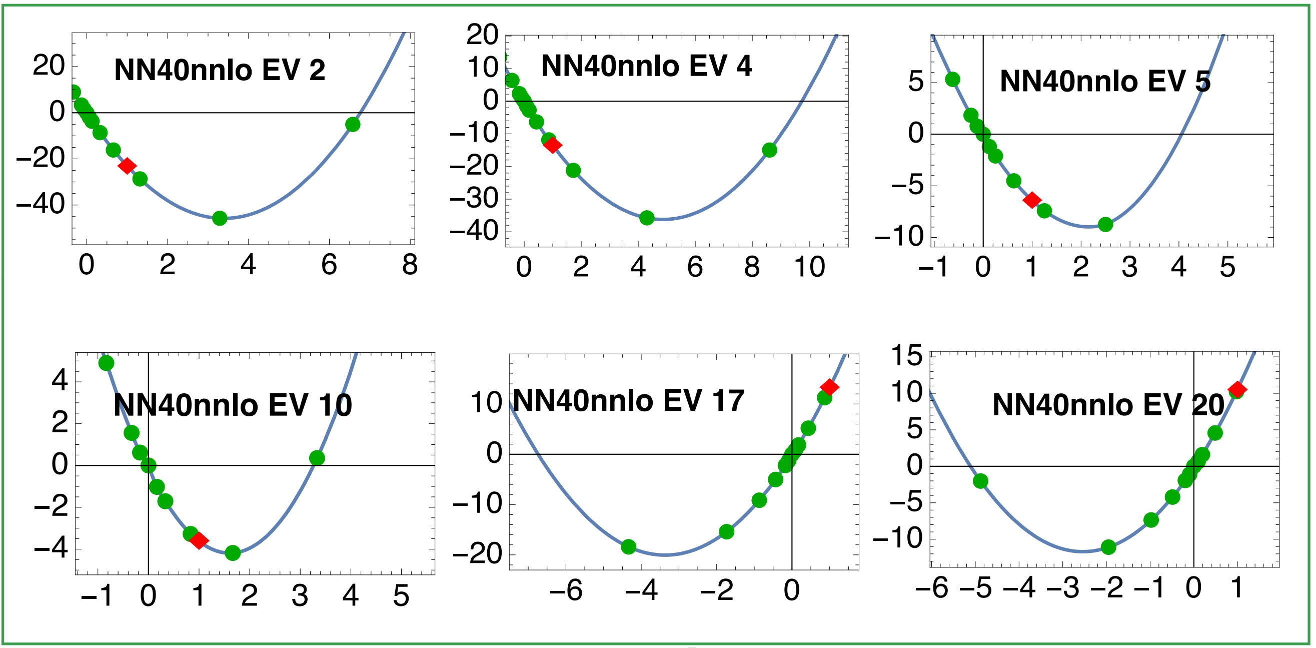
No fitting involved.



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

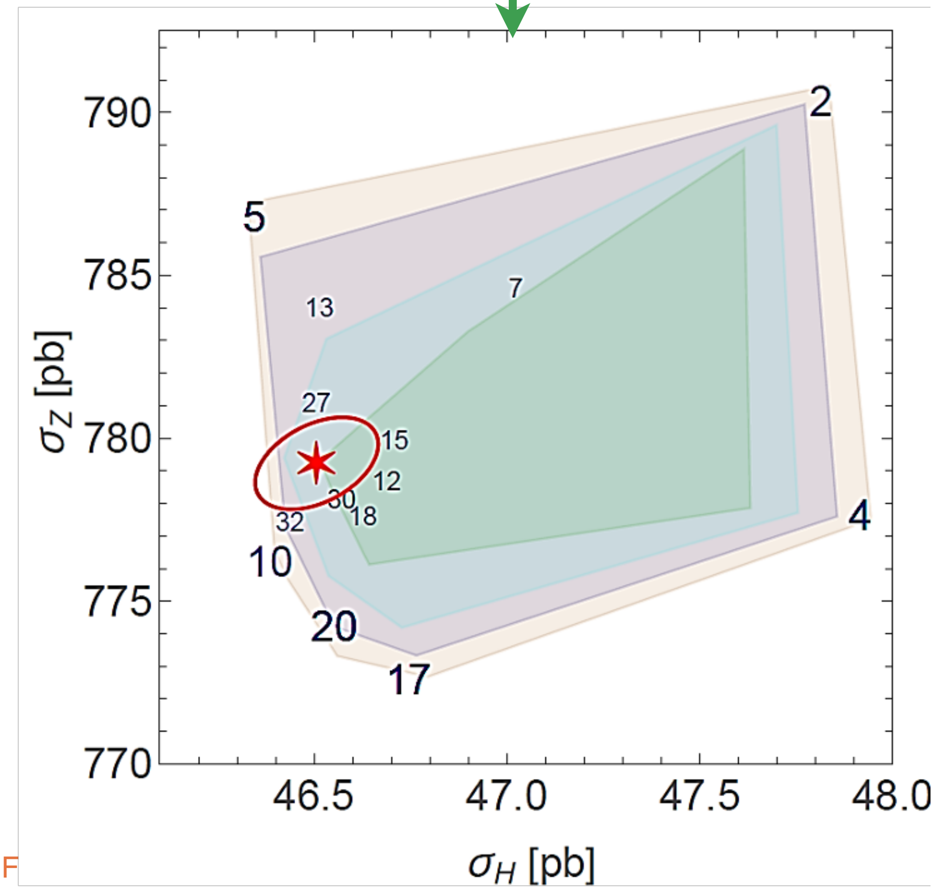
Step 3

Guidance from specific cross sections:
we identify 4-7 EV directions that give the largest displacements for a given $\Delta\chi^2$ per pair.



Large EV directions are shared among various pairs of cross sections.

Construct the convex hulls for $\Delta\chi^2 = +10, 0, -10, -20$ *w.r.t.* NNPDF4.0 replica 0 (red).



A hopscotch scan of LHC cross sections for NNPDF4.0 PDFs

Step 4

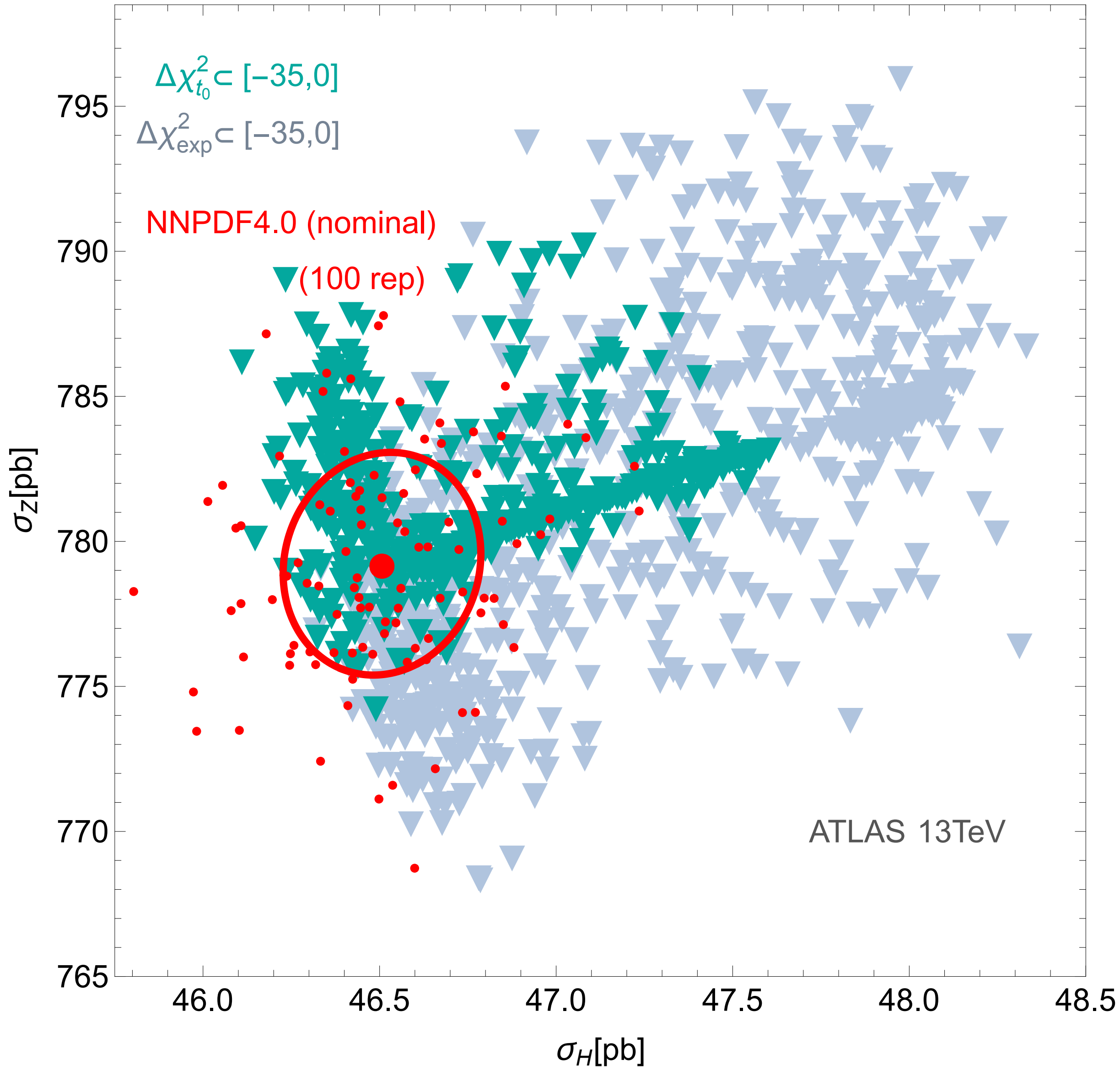
For each pair of cross sections, we generate 300 replicas by sampling uniformly along the “large” EV directions.

Sort the $n_{pairs} \times 300$ resulting replicas according to their $\Delta\chi^2$ *w.r.t.* to NN40 replica 0.

Hopscotch replicas are linear combinations of NNPDF4.0 Hessian EV.

Each of the solutions is an acceptable PDF set from the NNPDF4.0 fit.

High-density MC sampling of a span of a few EV directions that drive the specific PDF uncertainty.



Likelihoods in PDF analyses

$$\chi^2(a, \lambda^{exp}) = \sum_{i=1}^{N_{pts}} \left(\frac{T_i(a) - D_i + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha} \lambda_\alpha^{exp}}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \lambda_\alpha^{exp,2}$$

Simple algebraic eq. $\dfrac{d}{d\lambda^{exp}} \chi^2(a, \lambda^{exp}) = 0 \Rightarrow \bar{\lambda}^{exp}$

$$\chi^2(a) = \sum_{i=1}^{N_{pts}} \left(\frac{T_i(a) - D_i + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha} \bar{\lambda}_\alpha^{exp}}{\sigma_i} \right)^2 + \sum_{\alpha=1}^{N_\lambda} \bar{\lambda}_\alpha^{exp,2} + \chi^2_{prior}(a)$$

a is the vector of parameters of interest

β is the correlation matrix for nuisance parameters

Figures of merit in the NNPDF4.0 analysis I

1. χ^2 with respect to the central experimental values

$$\chi^2 = \sum_{i,j}^{N_{pt}} (T_i - D_i)(\mathrm{cov}^{-1})_{ij}(T_j - D_j)$$

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

D_i, T_i, s_i are the central data, theory, uncorrelated error
$\beta_{i,\alpha}$ is the correlation matrix for N_λ nuisance parameters.

Experiments publish $\sigma_{i,\alpha}$. To reconstruct $\beta_{i,\alpha}$, we need to decide on the normalizations X_i.

NNPDF4.0 use:

a. $X_i = D_i$: "**exp**erimental scheme"; can result in a bias
b. $X_i = \text{fixed } T_i$: "$\boldsymbol{t_0}$ scheme"; can result in a (different) bias

Figures of merit in the NNPDF4.0 analysis II

$$(\mathrm{cov})_{ij} \equiv s_i^2 \delta_{ij} + \sum_{\alpha=1}^{N_\lambda} \beta_{i,\alpha}\beta_{j,\alpha}, \qquad \beta_{i,\alpha} = \sigma_{i,\alpha} X_i,$$

NNPDF4.0 use:

a. $X_i = D_i$: **exp**erimental scheme; can result in a bias
b. $X_i =$ fixed T_i : $\boldsymbol{t_0}$ scheme; can result in a (different) bias

The conventions are neither complete nor unique. Ambiguity affects all groups. See Appendix in 1211.5142.

2. **NNPDF4.0 trains MC replicas with χ^2 for fluctuated D_i, t_0 scheme, and replica selection (prior) conditions:**

$$\text{Cost}=\chi^2_{t_0}(T_i, D_i^{fluctuated}) + \chi^2_{prior}$$

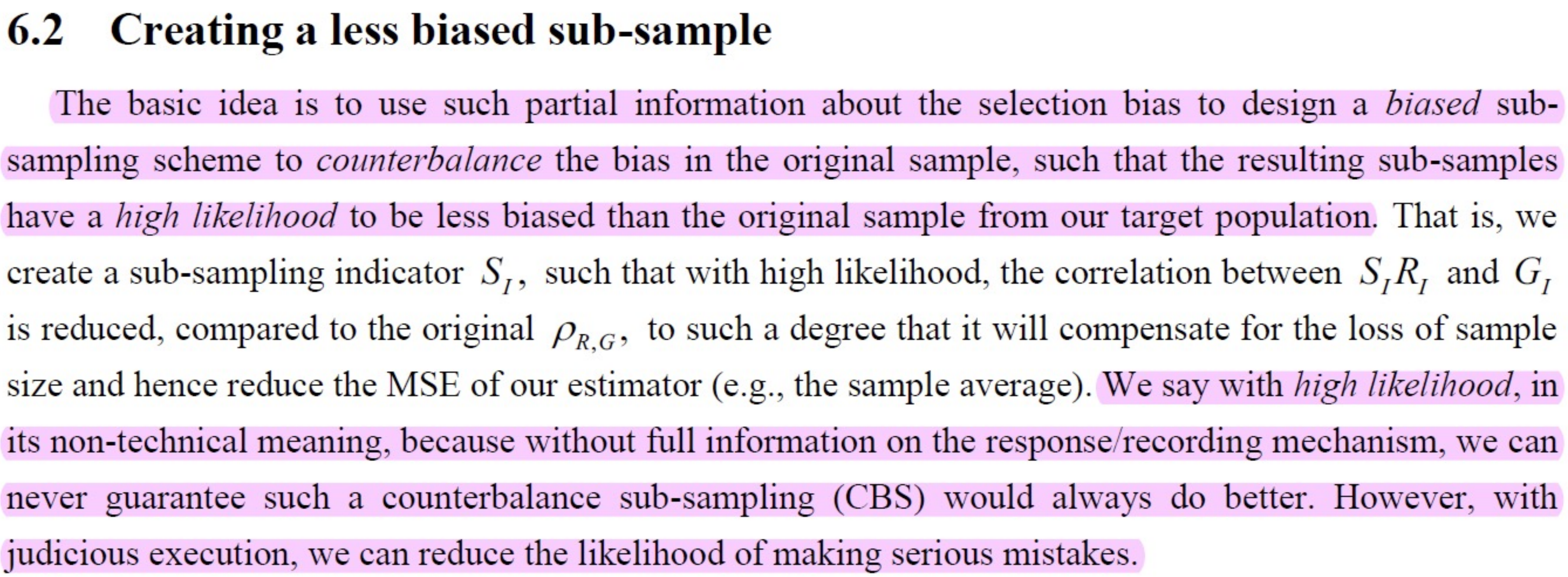
3. NNPDF4.0 quotes the final unfluctuated χ^2 in the “exp” scheme.

Experimental scheme:
$\chi^2_{tot}/N_{pt} = 1.160\,.$

t_0 scheme:
$\chi^2_{tot}/N_{pt} = 1.233\,.$

$\chi^2(\text{exp}) - \chi^2(t_0) = -340$ for 4618 data points

The hopscotch scan counterbalances the bias of the nominal replica ensemble

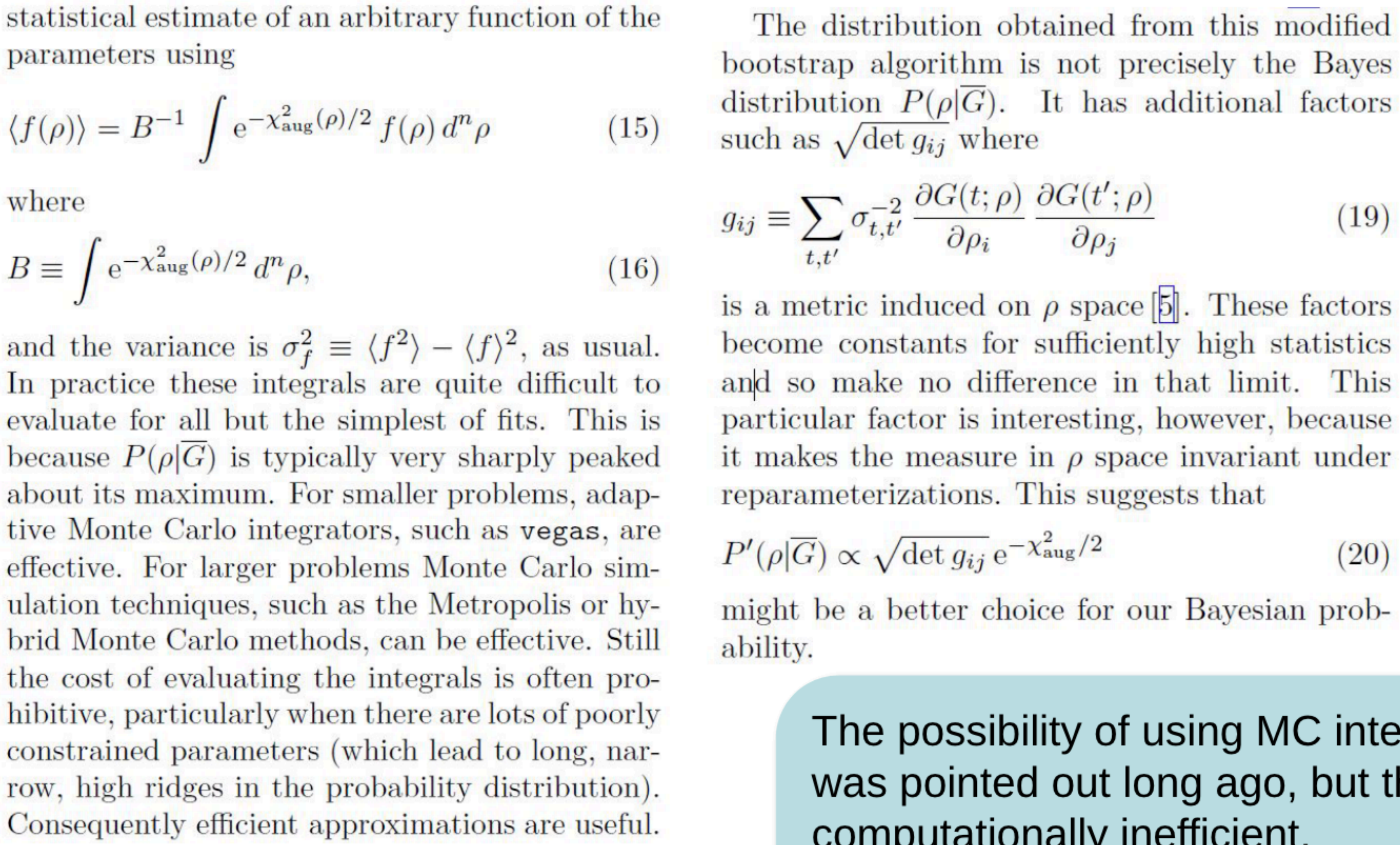


6.2 Creating a less biased sub-sample

The basic idea is to use such partial information about the selection bias to design a *biased* sub-sampling scheme to *counterbalance* the bias in the original sample, such that the resulting sub-samples have a *high likelihood* to be less biased than the original sample from our target population. That is, we create a sub-sampling indicator S_I, such that with high likelihood, the correlation between $S_I R_I$ and G_I is reduced, compared to the original $\rho_{R,G}$, to such a degree that it will compensate for the loss of sample size and hence reduce the MSE of our estimator (e.g., the sample average). We say with *high likelihood*, in its non-technical meaning, because without full information on the response/recording mechanism, we can never guarantee such a counterbalance sub-sampling (CBS) would always do better. However, with judicious execution, we can reduce the likelihood of making serious mistakes.

X.-L. Meng, Survey Methodology, Catalogue 12-001-X, vol. 48 (2022), #2

Priors and optimal sampling parameters



statistical estimate of an arbitrary function of the parameters using

$$\langle f(\rho)\rangle = B^{-1}\int \mathrm{e}^{-\chi^2_{\rm aug}(\rho)/2}\, f(\rho)\, d^n\rho \qquad (15)$$

where

$$B \equiv \int \mathrm{e}^{-\chi^2_{\rm aug}(\rho)/2}\, d^n\rho, \qquad (16)$$

and the variance is $\sigma_f^2 \equiv \langle f^2\rangle - \langle f\rangle^2$, as usual. In practice these integrals are quite difficult to evaluate for all but the simplest of fits. This is because $P(\rho|\overline{G})$ is typically very sharply peaked about its maximum. For smaller problems, adaptive Monte Carlo integrators, such as `vegas`, are effective. For larger problems Monte Carlo simulation techniques, such as the Metropolis or hybrid Monte Carlo methods, can be effective. Still the cost of evaluating the integrals is often prohibitive, particularly when there are lots of poorly constrained parameters (which lead to long, narrow, high ridges in the probability distribution). Consequently efficient approximations are useful.

The distribution obtained from this modified bootstrap algorithm is not precisely the Bayes distribution $P(\rho|\overline{G})$. It has additional factors such as $\sqrt{\det g_{ij}}$ where

$$g_{ij} \equiv \sum_{t,t'} \sigma^{-2}_{t,t'} \frac{\partial G(t;\rho)}{\partial \rho_i} \frac{\partial G(t';\rho)}{\partial \rho_j} \qquad (19)$$

is a metric induced on ρ space [5]. These factors become constants for sufficiently high statistics and so make no difference in that limit. This particular factor is interesting, however, because it makes the measure in ρ space invariant under reparameterizations. This suggests that

$$P'(\rho|\overline{G}) \propto \sqrt{\det g_{ij}}\, \mathrm{e}^{-\chi^2_{\rm aug}/2} \qquad (20)$$

might be a better choice for our Bayesian probability.

“Constrained curve fitting,” [Lepage et al., Nucl.Phys.B Proc.Suppl.106(2002) 12]

The possibility of using MC integration for expectation values was pointed out long ago, but the approach was deemed computationally inefficient.

Quasi-MC integration and dimensionality reduction may help, as well as parameter transformations to sample using a non-informative (e.g., Jeffrey’s) prior