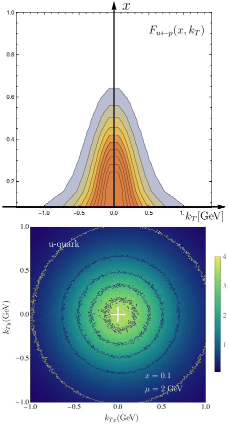
Hadron structure:
practice session

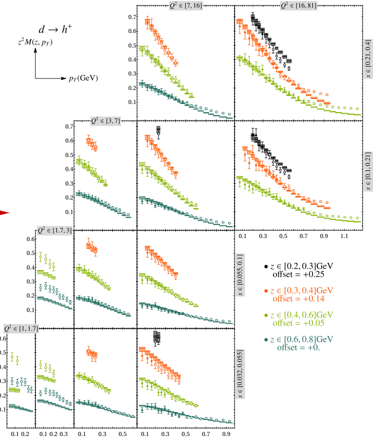
Alexey Vladimirov

MINISTERIO DE CIENCIA E INNOVACIÓN
Financiado por la Unión Europea
NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
AGENCIA ESTATAL DE INVESTIGACIÓN
UNIVERSITAS COMPLUTENSIS MATRITENSIS

AIMS of the practical session

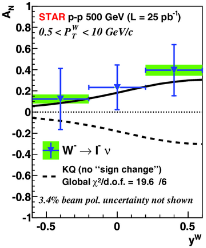
- How the hadron structure is determined? (practically)

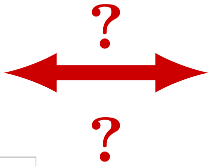


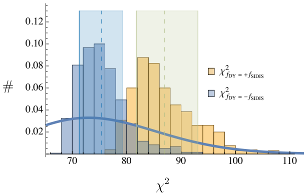


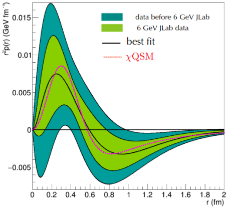
AIMS of the practical session

- Sources of uncertainties, and how to propagate them?



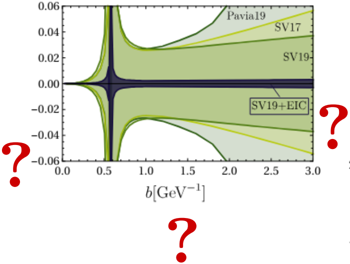


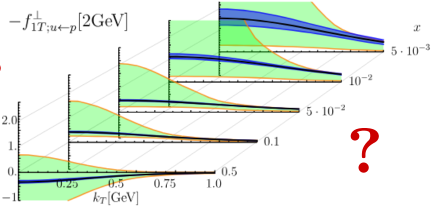




AIMS of the practical session

- Impact studies





Disclaimer

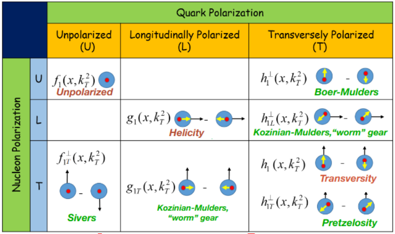
Instead of real formulas, we will use a toy-model and toy-data.
It is needed to save the time, because realistic computation could take hours/days.
Apart of it, the examples

Disclaimer

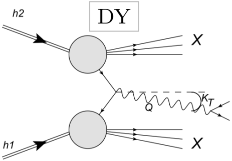
Instead of real formulas, we will use a toy-model and toy-data.
It is needed to save the time, because realistic computation could take hours/days.
Apart of it, the examples

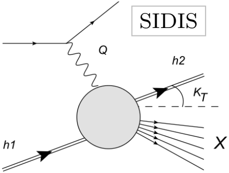
During EIC Yellow report preparation, there was a request to determine the requirements for EIC detector. As a test we selected the Sivers function...

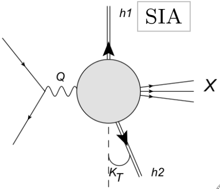
- The Sivers function was not known
- The event generators could not simulate such events

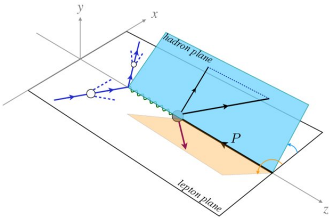


$$\frac{d\sigma}{dp_T} = \sigma_0 \int \frac{d^2b}{(2\pi)^2} e^{i(bp_T)} C\left(\frac{Q}{\mu}\right) F_1(x_1, b; \mu, \zeta) F_2(x_2, b; \mu, \bar{\zeta})$$









$$F_{UU,T} \sim f_1 \otimes D_1$$

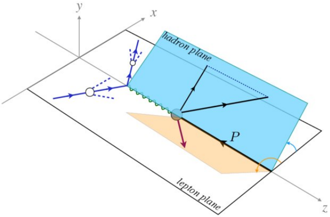
$$F_{UU}^{\cos 2\phi} \sim h_1^{\perp} \otimes H_1^{\perp}$$

$$F_{UL}^{\sin 2\phi} \sim h_{1L}^{\perp} \otimes H_1^{\perp}$$

$$F_{UT,T}^{\sin(\phi-\phi_S)} \sim \boxed{f_{1T}^{\perp}} \otimes D_1$$

...

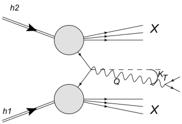
$$\begin{aligned}
&\frac{d\sigma}{dx\,dy\,d\psi\,dz\,d\phi_h\,dP_{h\perp}^2} = \\
&\frac{\alpha^2}{xyQ^2}\frac{y^2}{2\,(1-\varepsilon)}\left(1+\frac{\gamma^2}{2x}\right)\Big\{F_{UU,T}+\varepsilon F_{UU,L}+\sqrt{2\,\varepsilon(1+\varepsilon)}\,\cos\phi_h\,F_{UU}^{\cos\phi_h} \\
&\quad +\varepsilon\cos(2\phi_h)\,F_{UU}^{\cos 2\phi_h}+\lambda_e\sqrt{2\,\varepsilon(1-\varepsilon)}\,\sin\phi_h\,F_{LU}^{\sin\phi_h} \\
&\quad +S_{\parallel}\Big[\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin\phi_h\,F_{UL}^{\sin\phi_h}+\varepsilon\sin(2\phi_h)\,F_{UL}^{\sin 2\phi_h}\Big] \\
&+S_{\parallel}\lambda_e\Big[\sqrt{1-\varepsilon^2}\,F_{LL}+\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos\phi_h\,F_{LL}^{\cos\phi_h}\Big] \\
&+|\boldsymbol{S}_{\perp}|\Big[\sin(\phi_h-\phi_S)\left(F_{UT,T}^{\sin(\phi_h-\phi_S)}+\varepsilon\,F_{UT,L}^{\sin(\phi_h-\phi_S)}\right) \\
&\quad +\varepsilon\,\sin(\phi_h+\phi_S)\,F_{UT}^{\sin(\phi_h+\phi_S)}+\varepsilon\,\sin(3\phi_h-\phi_S)\,F_{UT}^{\sin(3\phi_h-\phi_S)} \\
&\quad +\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin\phi_S\,F_{UT}^{\sin\phi_S}+\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin(2\phi_h-\phi_S)\,F_{UT}^{\sin(2\phi_h-\phi_S)}\Big] \\
&+|\boldsymbol{S}_{\perp}|\lambda_e\Big[\sqrt{1-\varepsilon^2}\,\cos(\phi_h-\phi_S)\,F_{LT}^{\cos(\phi_h-\phi_S)}+\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos\phi_S\,F_{LT}^{\cos\phi_S} \\
&\quad +\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos(2\phi_h-\phi_S)\,F_{LT}^{\cos(2\phi_h-\phi_S)}\Big]\Big\},
\end{aligned}$$



$$\frac{d\sigma}{dx\,dy\,d\psi\,dz\,d\phi_h\,dP_{h\perp}^2} =$$
$$\frac{\alpha^2}{xyQ^2}\frac{y^2}{2(1-\varepsilon)}\left(1+\frac{\gamma^2}{2x}\right)\Big\{F_{UU,T}+\varepsilon F_{UU,L}+\sqrt{2\varepsilon(1+\varepsilon)}\cos\phi_h\,F_{UU}^{\cos\phi_h}$$
$$+\varepsilon\cos(2\phi_h)\,F_{UU}^{\cos 2\phi_h}+\lambda_e\sqrt{2\varepsilon(1-\varepsilon)}\sin\phi_h\,F_{LU}^{\sin\phi_h}$$
$$+S_\parallel\left[\sqrt{2\varepsilon(1+\varepsilon)}\sin\phi_h\,F_{UL}^{\sin\phi_h}+\varepsilon\sin(2\phi_h)\,F_{UL}^{\sin 2\phi_h}\right]$$
$$+S_\parallel\lambda_e\left[\sqrt{1-\varepsilon^2}\,F_{LL}+\sqrt{2\varepsilon(1-\varepsilon)}\cos\phi_h\,F_{LL}^{\cos\phi_h}\right]$$
$$+|\boldsymbol{S}_\perp|\Big[\sin(\phi_h-\phi_S)\left(F_{UT,T}^{\sin(\phi_h-\phi_S)}+\varepsilon F_{UT,L}^{\sin(\phi_h-\phi_S)}\right)$$
$$+\varepsilon\sin(\phi_h+\phi_S)\,F_{UT}^{\sin(\phi_h+\phi_S)}+\varepsilon\sin(3\phi_h-\phi_S)\,F_{UT}^{\sin(3\phi_h-\phi_S)}$$
$$+\sqrt{2\varepsilon(1+\varepsilon)}\sin\phi_S\,F_{UT}^{\sin\phi_S}+\sqrt{2\varepsilon(1+\varepsilon)}\sin(2\phi_h-\phi_S)\,F_{UT}^{\sin(2\phi_h-\phi_S)}\Big]$$
$$+|\boldsymbol{S}_\perp|\lambda_e\Big[\sqrt{1-\varepsilon^2}\cos(\phi_h-\phi_S)\,F_{LT}^{\cos(\phi_h-\phi_S)}+\sqrt{2\varepsilon(1-\varepsilon)}\cos\phi_S\,F_{LT}^{\cos\phi_S}$$
$$+\sqrt{2\varepsilon(1-\varepsilon)}\cos(2\phi_h-\phi_S)\,F_{LT}^{\cos(2\phi_h-\phi_S)}\Big]\Big\},$$

$$F_{UU,T}\sim f_1\otimes\boxed{D_1}$$
$$F_{UU}^{\cos 2\phi}\sim h_1^\perp\otimes H_1^\perp$$
$$F_{UL}^{\sin 2\phi}\sim h_{1L}^\perp\otimes H_1^\perp$$
$$F_{UT,T}^{\sin(\phi-\phi_S)}\sim\boxed{f_{1T}^\perp}\otimes\boxed{D_1}$$

...



$$d\sigma \simeq \boxed{f_1} \otimes \boxed{f_1}$$

$$F_{UU,T} \sim \boxed{f_1} \otimes \boxed{D_1}$$

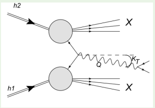
$$F_{UU}^{\cos 2\phi} \sim h_1^{\perp} \otimes H_1^{\perp}$$

$$F_{UL}^{\sin 2\phi} \sim h_{1L}^{\perp} \otimes H_1^{\perp}$$

$$F_{UT,T}^{\sin(\phi-\phi_S)} \sim \boxed{f_{1T}^{\perp}} \otimes \boxed{D_1}$$

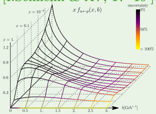
...

$$\begin{aligned}
&\frac{d\sigma}{dx\,dy\,d\psi\,dz\,d\phi_h\,dP_{h\perp}^2} = \\
&\frac{\alpha^2}{xyQ^2}\,\frac{y^2}{2\,(1-\varepsilon)}\left(1+\frac{\gamma^2}{2x}\right)\Big\{F_{UU,T}+\varepsilon\,F_{UU,L}+\sqrt{2\,\varepsilon(1+\varepsilon)}\,\cos\phi_h\,F_{UU}^{\cos\phi_h} \\
&\quad+\varepsilon\cos(2\phi_h)\,F_{UU}^{\cos 2\phi_h}+\lambda_e\,\sqrt{2\,\varepsilon(1-\varepsilon)}\,\sin\phi_h\,F_{LU}^{\sin\phi_h} \\
&\quad+S_{\parallel}\left[\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin\phi_h\,F_{UL}^{\sin\phi_h}+\varepsilon\sin(2\phi_h)\,F_{UL}^{\sin 2\phi_h}\right] \\
&+S_{\parallel}\lambda_e\left[\sqrt{1-\varepsilon^2}\,F_{LL}+\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos\phi_h\,F_{LL}^{\cos\phi_h}\right] \\
&+|\boldsymbol{S}_{\perp}|\Big[\sin(\phi_h-\phi_S)\left(F_{UT,T}^{\sin(\phi_h-\phi_S)}+\varepsilon\,F_{UT,L}^{\sin(\phi_h-\phi_S)}\right) \\
&\quad+\varepsilon\,\sin(\phi_h+\phi_S)\,F_{UT}^{\sin(\phi_h+\phi_S)}+\varepsilon\,\sin(3\phi_h-\phi_S)\,F_{UT}^{\sin(3\phi_h-\phi_S)} \\
&\quad+\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin\phi_S\,F_{UT}^{\sin\phi_S}+\sqrt{2\,\varepsilon(1+\varepsilon)}\,\sin(2\phi_h-\phi_S)\,F_{UT}^{\sin(2\phi_h-\phi_S)}\Big] \\
&+|\boldsymbol{S}_{\perp}|\lambda_e\Big[\sqrt{1-\varepsilon^2}\,\cos(\phi_h-\phi_S)\,F_{LT}^{\cos(\phi_h-\phi_S)}+\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos\phi_S\,F_{LT}^{\cos\phi_S} \\
&\quad+\sqrt{2\,\varepsilon(1-\varepsilon)}\,\cos(2\phi_h-\phi_S)\,F_{LT}^{\cos(2\phi_h-\phi_S)}\Big]\Big\},
\end{aligned}$$

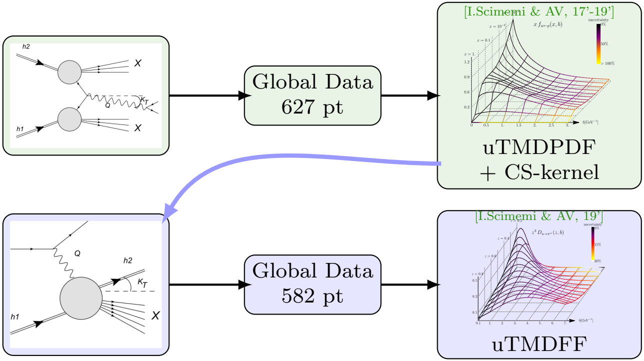


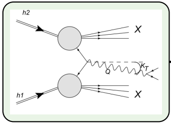
Global Data
627 pt

[I.Scimemi & AV, 17'-19']



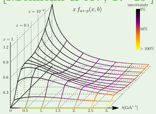
uTMDPDF
+ CS-kernel


h2
X
Q
K_T
h1
X
Global Data
627 pt
[I.Scimemi & AV, 17'-19']
uTMDPDF
+ CS-kernel
Q
h2
K_T
h1
X
Global Data
582 pt
[I.Scimemi & AV, 19']
uTMDFF

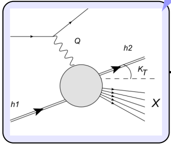


Global Data
627 pt

[I.Scimemi & AV, 17'-19']

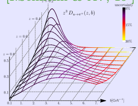


uTMDPDF
+ CS-kernel



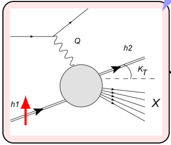
Global Data
582 pt

[I.Scimemi & AV, 19']



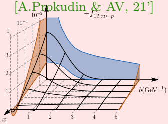
uTMDFF

+πTMD

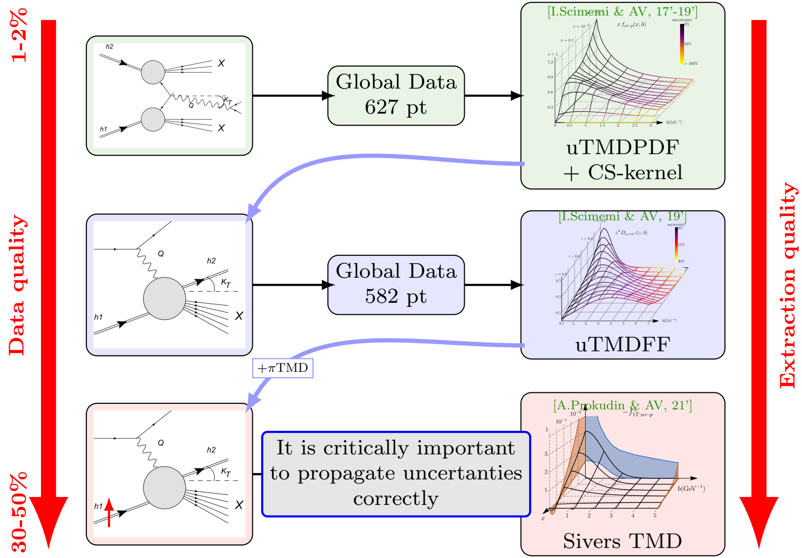


Global Data
76 pt

[A.Prokudin & AV, 21']

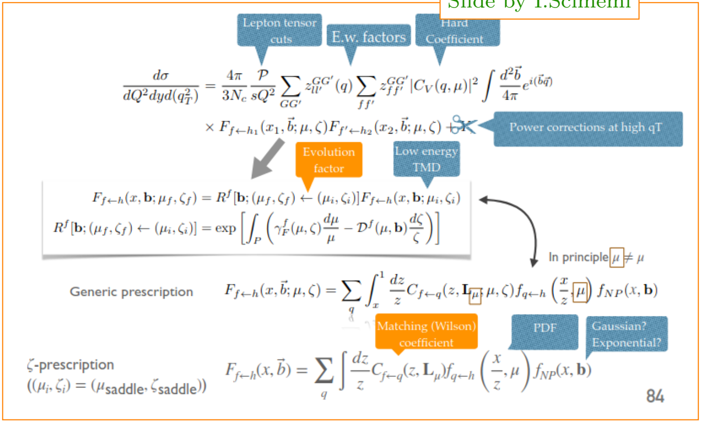


Sivers TMD


1-2%
Data quality
30-50%
h2
X
Q
k_T
h1
Global Data
627 pt
[I.Scimemi & AV, 17'-19']
uTMDPDF
+ CS-kernel
Global Data
582 pt
[I.Scimemi & AV, 19']
uTMDFF
+πTMD
It is critically important
to propagate uncertanties
correctly
[A.Prokudin & AV, 21']
Sivers TMD
Extraction quality

THEORY

Slide by I.Scimemi



If you want some real life example

- **artemide** – (standalone) fortran code with python interface
  https://github.com/VladimirovAlexey/artemide-public
- **nangaparbat** – C++ code (requires quite a lot of extra packages)
  https://github.com/MapCollaboration/NangaParbat
- ...

TMD factorization

$$\frac{d\sigma}{dQdq_Tdxdz} \simeq \sum_{q,\bar{q}} e_q^2 C(Q) \int_0^\infty db\, b J_0(bq_T) \left(\frac{Q^2}{\zeta_Q(b)}\right)^{-2\mathcal{D}(b,Q)} F_q(x,b) F_{\bar{q}}(z,b) \tag{1}$$

- J_0=Bessel function (very expensive integral)
- $\zeta_Q(b)$ = rather complicated function
- C = hard coefficient function
- At $b \to 0$ all functions are known (perturbatively via PDFs), Q dependence is known

Toy model

$$\frac{d\sigma}{dQdq_Tdxdz} = (4\pi)^2 \int_0^\infty db\, b e^{-(bq)_T} \cos(bq_T) \left(\frac{Q^2}{(2\text{GeV})^2}\right)^{-2\mathcal{D}(b,Q)} F_1(x,b) F_2(z,b) \tag{2}$$

- generally has very similar behavior as TMD factorization.
- Can be computed with usual integration routine in a finite range ($b \in [0, 25/q_T]$ provides 10^{-6}-precision)

Suggested models (1)

Collins-Soper kernel

$$\mathcal{D}(b,Q) = \underbrace{0.004 \ln\left(\frac{\mathbf{b}^2 Q^2}{4e^{-2\gamma_E}}\right) - 0.001}_{\sim\text{perturbation theory}} + D_{\text{NP}}(b) \tag{3}$$

- $\lim_{b\to 0} D_{\text{NP}}(b) = 0 + b^2 ...$ $\qquad$ $\lim_{b\to\infty} D_{\text{NP}}(b) > 0$
- Example: $D_{\text{NP}}(b) = \alpha_1 b^2$, or $D_{\text{NP}}(b) = \dfrac{\alpha_1 b^2}{\sqrt{1+\alpha_2 b^2}}$

Suggested models (2)

Unpolarized TMDPDF

$$f_1(x,b) = f(x) f_{\rm NP}(x,b) \tag{4}$$

- $f(x)$= collinear PDF
- $f_{\rm NP}(x,b)$ is any function such that

$$\lim_{b\to 0} f_{\rm NP}(x,b) = 1 + b^2..., \qquad \lim_{b\to\infty} f_{\rm NP}(x,b) = 0$$

- Example

$$f(x,b) = \frac{1}{\cosh[(x\beta_1 + (1-x)\beta_2)b]}, \qquad f(x,b) = e^{-\beta_1 b^2}(1 + x\beta_2 b^2)$$

Suggested models (3)

Unpolarized TMDPDF

$$f_1(x,b) = f(x) f_{\rm NP}(x,b) \quad (5)$$

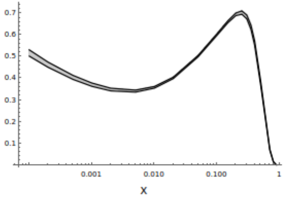
- PDF ($\simeq$ HERAPDF2.0) with a single flavor

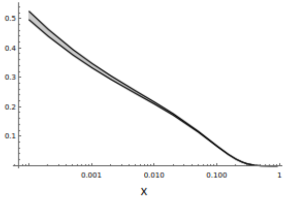
$$\boxed{f(x) = Ax^B(1-x)^C(1+Ex^2) + \bar{A}x^{\bar{B}}(1-x)^{\bar{C}},} \qquad \boxed{\bar{f}(x) = \bar{A}x^{\bar{B}}(1-x)^{\bar{C}}}$$

- All parameters **uncorrelated**

$$A = 4.07 \pm 0.002, \quad B = 0.714 \pm 0.002, \quad C = 4.84 \pm 0.03, \quad E = 13.4 \pm 0.1,$$

$$\bar{A} = 0.105 \pm 0.001, \quad \bar{B} = -0.172 \pm 0.003, \quad \bar{C} = 8.06 \pm 0.04$$





Realization notes:

- Specify your model and number of parameters

```
#### Write your model for CS kernel
Nparam_CS=2  ##### specify number of parameters in it

def D_NP(b,CSparam):
    return CSparam[0]*b*b/np.sqrt(1.+CSparam[1]*b*b)
```

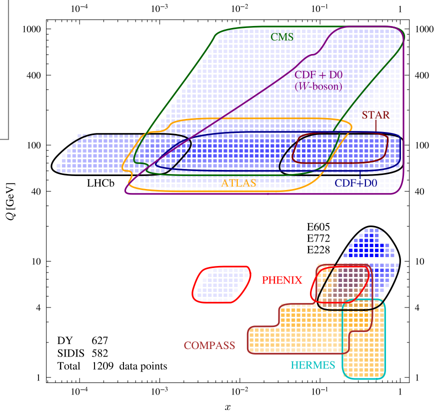
```
#### Write your model for TMDPDF
Nparam_TMD=2  ##### specify number of parameters in it

def f_NP1(x,b,TMDparam):
    return 1/np.cosh(TMDparam[0]*b)
def f_NP2(x,b,TMDparam):
    return 1/np.cosh(TMDparam[1]*b)
```

- Call `xSec(Q,x1,x2,qT,p1,p2)` or `xSec_bar(Q,x1,x2,qT,p1,p2)` to compute cross-section
  - Q, x1, x2, qT are kinematic parameters
  - p1=list of NP parameter [CS,TMD]
  - p2=list of parameters for PDF (use `PDF_parameters_central` for default value)

DATA

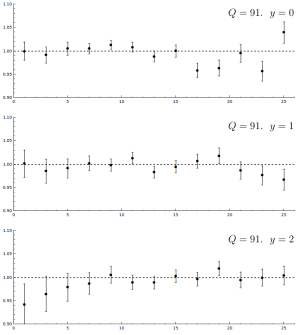
ART25

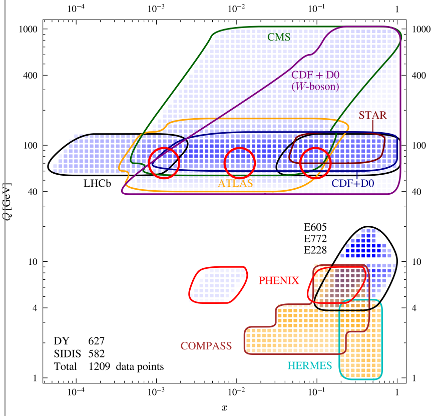
- Low ($Q = 2$ GeV) and high ($Q = 1$ TeV) energies
- Different flavor content pp, $p\bar{p}$, $p\pi^{\pm}$, $pK^{\pm}$...
- Different x's



"ATLAS"

- $Q = 91\text{GeV}$
- Presice ($\sim 1-3\%$)
- 3 cases of $x_{1,2}$
- **LABEL**: A0, A1, A2



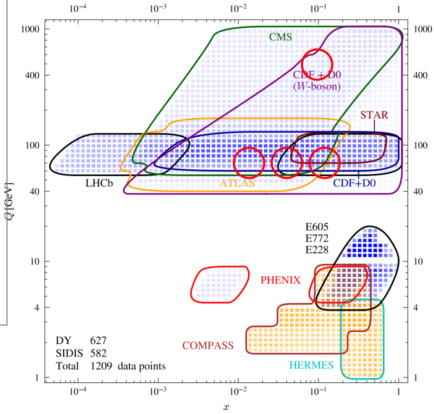


"CMS"

- $Q = 500\text{GeV}$
- Moderate uncertanty ($\sim 5-10\%$)
- **LABEL**: CMS

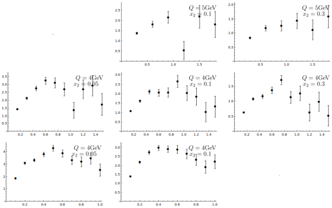
"Tevatron"

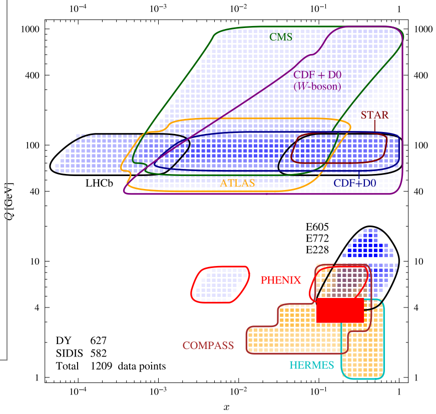
- $Q = 91\text{GeV}$
- Moderate uncertanty ($\sim 5-7\%$)
- $p\bar{p}$ (!)
- 2 cases of $x_{1,2}$
- **LABEL**: T0, T1



"COMPASS"

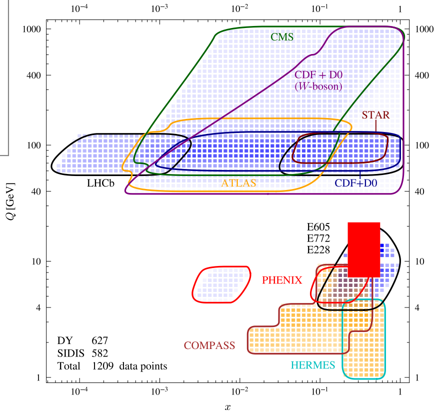
- $Q = 3, 4, 5$GeV, different x's
- pp and $p\bar{p}$ cases
- uncertanty ($\sim 2 - 50\%$)
- **LABEL**: "COM+", "COM-"





"Fixed Target"

- $Q = 8 - 15$GeV, different x's
- pp
- Large uncertanty ($\sim 10 - 50\%$) $+ 25\%$ normalization
- **LABEL**: "FT"

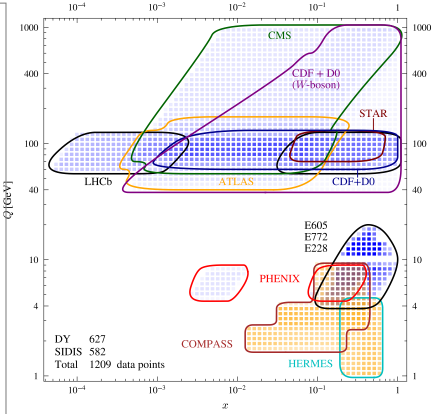


Total:

- 276 points
- Computation of cross-section on "avarage" compute $\sim 0.3 - 0.6$ sec.

Practical notes:

- Jupiter notebook reads data-files and creates `data["name"]`-lists
- All data presented as `[Q,x1,x2,qT,xSec,unc1,unc2]` where "unc" are two uncorrelated uncertanties
- Each set has its own normalization (due to luminocity) uncertanty (in %) it can be read in `lumUNC["name"]`



How to compare theory with data?

Slide by E.Nocera

Fit quality

1. Define the fit quality (the χ^2 function)

$$\chi^2 = \sum_{i,j}^{\mathrm{N_{dat}}} \left(T_i[\{\mathbf{a}\}] - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j[\{\mathbf{a}\}] - D_j\right)$$

with the experimental covariance matrix

$$(\mathrm{cov})_{ij} = \delta_{ij} s_i^2 + \left(\sum_{\alpha}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \sum_{\alpha}^{N_{\mathcal{L}}} \sigma_{i,\alpha}^{(\mathcal{L})} \sigma_{j,\alpha}^{(\mathcal{L})} \right) D_i D_j$$

s_i are N_{dat} uncorrelated uncertainties (statistic + uncorrealted systematic ucnertainties)
$\sigma_{i,\alpha}^{(c)}$ are $N_{\mathrm{dat}} \times N_c$ additive correlated uncertainties
$\sigma_{i,\alpha}^{(\mathcal{L})}$ are $N_{\mathrm{dat}} \times N_{\mathcal{L}}$ multiplicative uncertainties

How to compare theory with data?

Slide by E.Nocera

Fit quality

1. Define the fit quality (the χ^2 function)

$$\chi^2 = \sum_{i,j}^{\mathrm{N_{dat}}} \left(T_i[\{\mathbf{a}\}] - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j[\{\mathbf{a}\}] - D_j\right)$$

with the experimental covariance matrix

$$(\mathrm{cov})_{ij} = \delta_{ij} s_i^2 + \left(\sum_{\alpha}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \sum_{\alpha}^{N_{\mathcal{L}}} \sigma_{i,\alpha}^{(\mathcal{L})} \sigma_{j,\alpha}^{(\mathcal{L})} \right) D_i D_j$$

s_i are N_{dat} uncorrelated uncertainties (statistic + uncorrealted systematic ucnertainties)
$\sigma_{i,\alpha}^{(c)}$ are $N_{\mathrm{dat}} \times N_c$ additive correlated uncertainties
$\sigma_{i,\alpha}^{(\mathcal{L})}$ are $N_{\mathrm{dat}} \times N_{\mathcal{L}}$ multiplicative uncertainties

All uncorrelated uncertanties as "Gaussian noise"

$$\sqrt{u_1^2 + u_2^2 + \ldots}$$

statistical, uncorelated effects, etc.

How to compare theory with data?

Slide by E.Nocera

Fit quality

1. Define the fit quality (the χ^2 function)

$$\chi^2 = \sum_{i,j}^{\mathrm{N_{dat}}} \left(T_i[\{\mathbf{a}\}] - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j[\{\mathbf{a}\}] - D_j\right)$$

with the experimental covariance matrix

$$(\mathrm{cov})_{ij} = \delta_{ij} s_i^2 + \left(\sum_{\alpha}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \sum_{\alpha}^{N_{\mathcal{L}}} \sigma_{i,\alpha}^{(\mathcal{L})} \sigma_{j,\alpha}^{(\mathcal{L})} \right) D_i D_j$$

s_i are N_{dat} uncorrelated uncertainties (statistic + uncorrealted systematic ucnertainties)
$\sigma_{i,\alpha}^{(c)}$ are $N_{\mathrm{dat}} \times N_c$ additive correlated uncertainties
$\sigma_{i,\alpha}^{(\mathcal{L})}$ are $N_{\mathrm{dat}} \times N_{\mathcal{L}}$ multiplicative uncertainties

All additive correlated uncertanties in %
features of detector, etc.
absent in our toy case

How to compare theory with data?

Slide by E.Nocera

Fit quality

1. Define the fit quality (the χ^2 function)

$$\chi^2 = \sum_{i,j}^{\mathrm{N_{dat}}} \left(T_i[\{\mathbf{a}\}] - D_i\right) \left(\mathrm{cov}^{-1}\right)_{ij} \left(T_j[\{\mathbf{a}\}] - D_j\right)$$

with the experimental covariance matrix

$$(\mathrm{cov})_{ij} = \delta_{ij} s_i^2 + \left(\sum_{\alpha}^{N_c} \sigma_{i,\alpha}^{(c)} \sigma_{j,\alpha}^{(c)} + \sum_{\alpha}^{N_{\mathcal{L}}} \sigma_{i,\alpha}^{(\mathcal{L})} \sigma_{j,\alpha}^{(\mathcal{L})} \right) D_i D_j$$

s_i are N_{dat} uncorrelated uncertainties (statistic + uncorrealted systematic ucnertainties)
$\sigma_{i,\alpha}^{(c)}$ are $N_{\mathrm{dat}} \times N_c$ additive correlated uncertainties
$\sigma_{i,\alpha}^{(\mathcal{L})}$ are $N_{\mathrm{dat}} \times N_{\mathcal{L}}$ multiplicative uncertainties

All multiplicative correlated uncertanties in %
luminocity uncertanty, depolarization, etc..

How to compare theory with data?

Slide by E.Nocera

Fit quality

1. Define the fit quality (the χ^2 function)

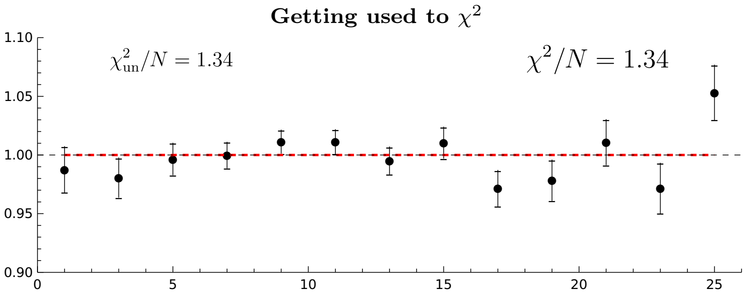
$$\chi^2 = \sum_{i,j}^{N_{\rm dat}} (T_i[\{\mathbf{a}\}] - D_i)\,(\mathrm{cov}^{-1})_{ij}\,(T_j[\{\mathbf{a}\}] - D_j)$$

Realisation note:

- `TheoryPrediction(Name,p1,p2)` computes the theory prediction for data set with "Name" and NP parameters p1 (TMD) and p2 (PDF)
- `chi2(Name,p1,p2)` computes χ^2 with same input
- `chi2Total(p1,p2)` computes χ^2 for the main data set with same input

$\sigma_{i,\alpha}$ are $N_{\rm dat} \times N_c$ additive correlated uncertainties

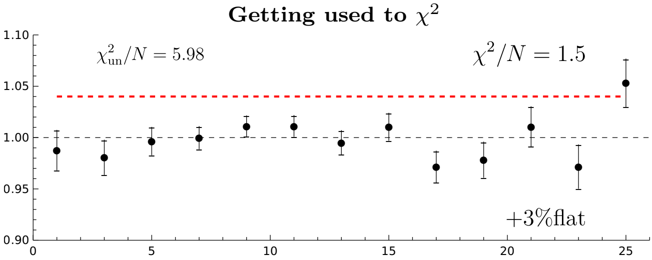
$\sigma^{(\mathcal{L})}_{i,\alpha}$ are $N_{\rm dat} \times N_{\mathcal{L}}$ multiplicative uncertainties



- $\chi^2/N \approx 1$ implies that agreement between data and theory approximately within 1σ of each point. Which is correct for randomly-generated data.
- $\chi^2/N \gg 1$ Model does not describe the data
- $\chi^2/N \ll 1$ Data is "too smooth" or model is "overfitted"

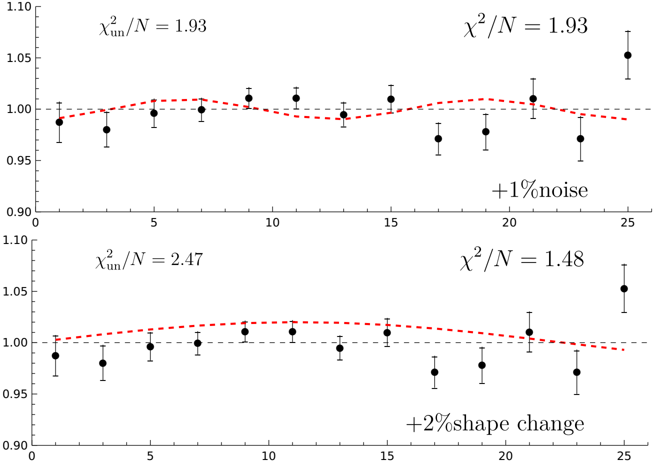
Excellent (short and concrete) introduction to χ^2

https://www.nevis.columbia.edu/ seligman/root-class/html/appendix/statistics/index.html



- Uncorrelated and correlated uncertainties weights “deviations” from data differently

Getting used to χ^2



- Uncorrelated and correlated uncertainties weights "deviations" from data differently

The minimization algorithm is your best friend/worst enemy

Slide by E.Nocera

Optimisation should minimise the noise in the χ^2 driven by noisy experimental data

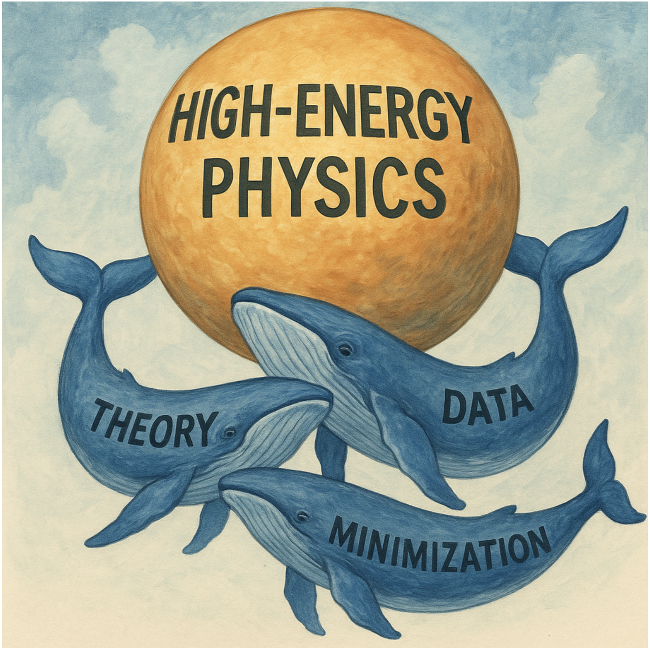
Additional complications in case of a redundant parametrisation (huge parameter space)

1. need to explore the parameter space as uniformly as possible (in order to avoid stopping the fit in a local minimum)
2. need for a computationally efficient minimisation (non-trivial relationship between FFs and observables via convolution)
3. need to define a criterion for minimisation stopping (avoid learning statistical fluctuations of the data)

Alternative algorithms:
genetic algorithms, adaptive algorithms, ...

Practical notes

- In notebook, there are several examples of minimization routines
- For our simple case the built-in (`scipy.optimize.minimize`) works well, but for full-scale fits (with 10+ parameters) minimization is a principal issue.
- Example task requires $\sim 1-2$min to be minimized ($\sim 10-20$sec, if close to minimum)


HIGH-ENERGY PHYSICS
THEORY
DATA
MINIMIZATION

Task 1:

- Invent a model
- Fit the "unpolarized data"

Task 2 (hard):

- Invent a model for "Sivers" function
- Fit the "polarized data"
  - The unpolarized TMDPDF is determined in **Task 1**

Estimation of uncertainty

Slide by E.Nocera

The Hessian method: general strategy

1. Expand the χ^2 about its global minimum at first (nontrivial) order

$$\chi^2\{\mathbf{a}\} \approx \chi^2\{\mathbf{a_0}\} + \delta a^i H_{ij} \delta a^j, \qquad H_{ij} = \frac{1}{2} \left. \frac{\partial^2 \chi^2\{\mathbf{a}\}}{\partial a_i \partial a_j} \right|_{\{\mathbf{a}\}=\{\mathbf{a_0}\}}$$

2. Assume linear error propagation for any observable $\mathcal{O}$ depending on $\{\mathbf{a}\}$

$$\langle \mathcal{O}\{\mathbf{a}\} \rangle \approx \mathcal{O}\{\mathbf{a_0}\} + a_i \left. \frac{\partial \mathcal{O}\{\mathbf{a}\}}{\partial a_i} \right|_{\{\mathbf{a}\}=\{\mathbf{a_0}\}} \qquad \sigma_{\mathcal{O}\{\mathbf{a}\}} \approx \sigma_{ij} \left. \frac{\partial \mathcal{O}\{\mathbf{a}\}}{\partial a_i} \frac{\partial \mathcal{O}\{\mathbf{a}\}}{\partial a_j} \right|_{\{\mathbf{a}\}=\{\mathbf{a_0}\}}$$

3. Determine σ_{ij} from H_{ij} from maximum likelihood (under Gaussian hypothesis)

$$\sigma_{ij}^{-1} = \left. \frac{\partial^2 \chi^2\{\mathbf{a}\}}{\partial a_i \partial a_j} \right|_{\{\mathbf{a}\}=\{\mathbf{a_0}\}} = H_{ij}$$

4. A C.L. about the best fit is obtained as the volume (in parameter space) about $\chi^2\{\mathbf{a_0}\}$ that corresponds to a fixed increase of the χ^2; for Gaussian uncertainties:

$$68\%\ \text{C.L.} \Longleftrightarrow \Delta\chi^2 = \chi^2\{\mathbf{a}\} - \chi^2\{\mathbf{a_0}\} = 1$$

This is important part of data-analyses, but it requires more programming + time.
We skip it here.

Estimation of uncertainty

Slide by E.Nocera

The Monte Carlo method: general strategy

1. Generate (art) replicas of (exp) data according to the distribution

$$\mathcal{O}_i^{(art)(k)} = \mathcal{O}_i^{(exp)} + r_i^{(k)}\sigma_{\mathcal{O}_i}\,, \qquad i = 1, \dots N_{\text{dat}}\,, \qquad k = 1, \dots, N_{\text{rep}}$$

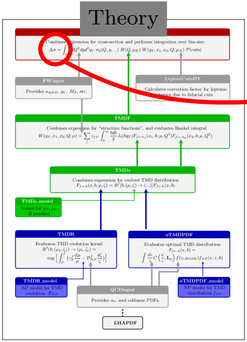
   where $r_i^{(k)}$ are (Gaussianly distributed) random numbers for each k-th replica ($r_i^{(k)}$ can be generated with any distribution, not neccesarily Gaussian)

2. Perform a fit for each replica $k = 1, \dots, N_{\text{rep}}$
3. Compact computation of observables and their uncertainties (PDF replicas are equally probable members of a statistical ensemble)

$$\langle \mathcal{O}[f(x,Q^2)]\rangle = \frac{1}{N_{\text{rep}}}\sum_{k=1}^{N_{\text{rep}}} \mathcal{O}[f^{(k)}(x,Q^2)]$$

$$\sigma_{\mathcal{O}}[f(x,Q^2)] = \left[\frac{1}{N_{\text{rep}}-1}\sum_{k=1}^{N_{\text{rep}}}\left(\mathcal{O}[f^{(k)}(x,Q^2)] - \langle \mathcal{O}[f(x,Q^2)]\rangle\right)^2\right]^{1/2}$$

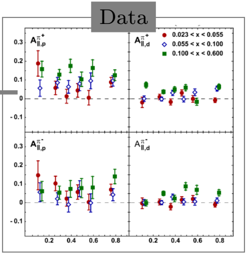
$\Rightarrow$ **no need to rely on linear approximation**
$\Rightarrow$ **computational expensive: need to perform N_{rep} fits instead of one**



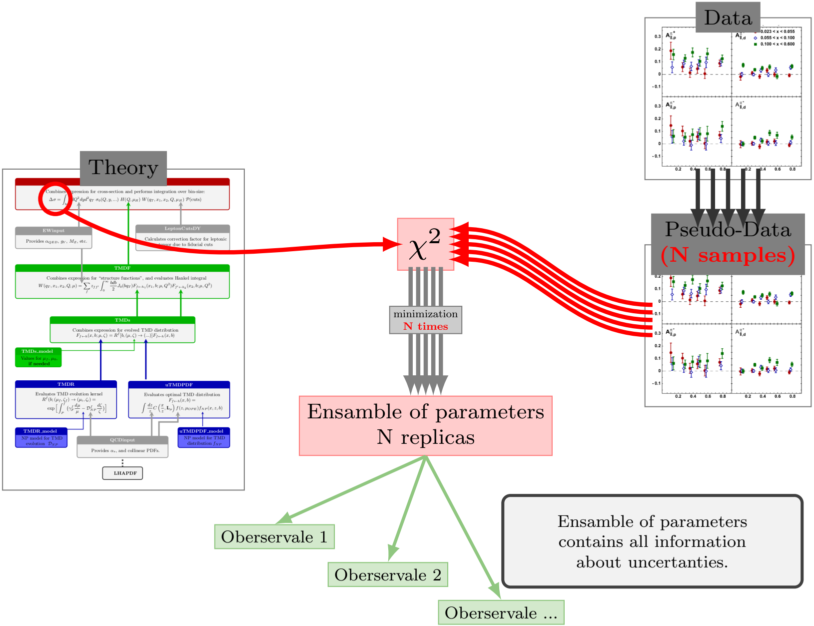
χ^2

minimization

Values of parameters (central)

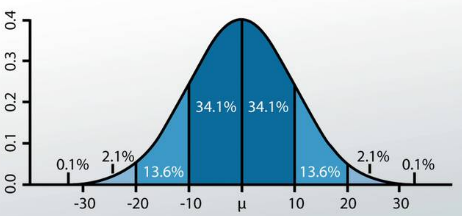


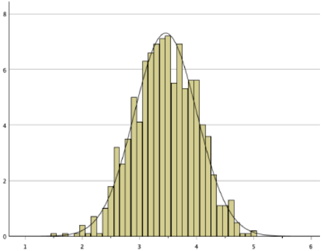
Central value is not very usefull, because it is just a single representative of a random distribution (although it should be close to the mean value)


Theory
Data
Pseudo-Data
(N samples)
χ^2
minimization
N times
Ensamble of parameters
N replicas
Oberservale 1
Oberservale 2
Oberservale ...
Ensamble of parameters
contains all information
about uncertanties.

Mean value, uncertanty, etc.

- Mean, uncertanty, and other integrated quontities are never exact. Only complete ensamble is incorporate all information.
- It is especially correct for multi-parameteric observables.
- Ideal case = Gaussian distribution

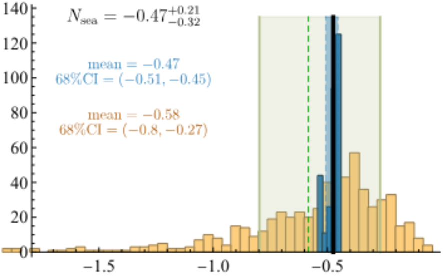




- Std $\approx 68\%$ of events.

Mean value, uncertanty, etc.

- Often one gets something like that



- **Problem 1:** Not Gaussian, not symmetric
- **Solution:** Generalize concepts (e.g. 68%C.I. is the part of events if one drops 16% events from the sides.)
- **Problem 2:** Descrite, and noisy
- **Solution:** Bootstrap method: *A. Davison and D. Hinkley, Bootstrap Methods and their Application*
  - Make a random sub-set (with repetitions) of the set
  - Compute observable
  - Do it MANY times
  - Avarage result

END OF PART I

Task to be done before tomorrow

- Complete the fit (with uncertanties) of your model
  Possible speed-up hint: To speed-up the computation you can drop some of data-sets. I suggest to drop "FT", "CMS", (if does not help) "A2", "A1", ... In the order of importantce, but do not drop too much!
- Compute the mean and uncertanties on the obtained parameters (are they symmetric?)
- **(Difficult)** Do the same for Sivers function [the toy-theory and toy-data are introduced in the file `Sivers.ipynb`