Bayesian reweighting for PDFs

(arXiv:1310.1089)

Nobuo Sato
Florida State University

In collaboration with:
J. Owens
H. Prosper

Background/motivation:

- The reweighting method allows us to modify PDFs to include new data without performing a global fit.
- The technique was proposed by Giele and Keller (hep-ph/9803393) and later developed by the NNPDF collaboration (hep-ph/1012.0836, hep-ph/0912.2276).
- If the theoretical description of the new data is time consuming for global fits, the reweighting is an efficient alternative.
- It can be seen as a complementary tool for global fits.

Notation:

- pdf : probability density function.
- PDF : parton distribution function.

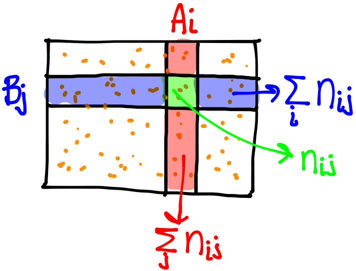
Outline:

- The reweighting technique.
- A recipe to reweight PDFs.
- Application of the reweighting: single inclusive direct photon data from fixed target experiments.

The reweighting technique

Bayesian statistics in a nutshell:

- Consider two observables A and B and a sample of N data points $\{A_i, B_j\}$.



$$P(A_i, B_j) = \frac{n_{ij}}{N} \quad (1)$$

$$P(A_i) = \frac{\sum_j n_{ij}}{N} \quad (2)$$

$$P(A_i|B_j) = \frac{n_{ij}}{\sum_i n_{ij}} \quad (3)$$

$$= \frac{n_{ij}}{N} \frac{N}{\sum_i n_{ij}} \quad (4)$$

$$= \frac{P(A_i, B_j)}{P(B_j)} \quad (5)$$

$$P(A_i|B_j)P(B_j) = P(A_i, B_j) \quad (6)$$

The Bayes theorem:

$$P(B|A) = \frac{P(A|B)}{P(A)} P(B) \quad (7)$$

$= $ *probability of B given A*

The reweighting $\equiv$ Bayes theorem

- Consider a model with parameters $\vec{\alpha}$ that describes some observable. (e.g cross sections as a function of p_T distribution)
- Using some data (labeled as D_{old}), we fit the parameters $\vec{\alpha}$.
- The uncertainties of the fit and its central values gives an estimate of the parent distribution (a pdf) for $\vec{\alpha}$: $\mathcal{P}(\vec{\alpha})$
- With a new evidence D_{new} the Bayes theorem states that:

$$\mathcal{P}(\vec{\alpha}|D_{new}) = \frac{\mathcal{P}(D_{new}|\vec{\alpha})}{\mathcal{P}(D_{new})}\mathcal{P}(\vec{\alpha}) \tag{8}$$

$$posterior = likelihood \quad \times \quad prior$$

- The *posterior* depends on how we quantify *Likelihood*.

How to construct the *Likelihood* $\propto \mathcal{P}(D_{new}|\vec{\alpha})$?

- Suppose that the new data consist of n data points arranged as a vector $\vec{y}$ with covariance matrix Σ. (for simplicity lets consider only uncorrelated errors.)

- Using the prior pdf $\mathcal{P}(\vec{\alpha})$ we compute the n predictions $\vec{t}$ for the new data.

- Assuming Gaussian statistics we can write

$$\begin{aligned}\mathcal{P}(\vec{y}|\vec{\alpha})\, d^n y &= \prod_j \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2}\left(\frac{y_j - t_j}{\sigma_j^2}\right)^2} dy_j \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\chi^2(\vec{y},\vec{t})}\, d^n y, \qquad (9)\end{aligned}$$

How to construct the *Likelihood* $\propto \mathcal{P}(D_{new}|\vec{\alpha})$?

$$\mathcal{P}(\vec{y}|\vec{\alpha})\, d^n y = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\chi^2(\vec{y},\vec{t})}\, d^n y \tag{10}$$

- Notice that we can write

$$d^n y = \chi^{n-1}\ d\chi\ d\Omega_{n-1} \tag{11}$$

- Alternatively, the probability of the new data to be confined in a differential shell χ to $\chi + d\chi$ is given by

$$\mathcal{P}(\chi|\vec{\alpha})\, d\chi = \frac{1}{2^{n/2-1}\Gamma(n/2)} \chi^{n-1} e^{-\frac{1}{2}\chi^2}\, d\chi, \tag{12}$$

- By construction $\mathcal{P}(\chi|\vec{\alpha})$ contains less information than $\mathcal{P}(\vec{y}|\vec{\alpha})$.

The recipe for reweighting

- Suppose we have at our disposal a model with fitted parameters $\vec{\alpha}$ with its uncertainties ($\equiv \mathcal{P}(\vec{\alpha})$).

- Suppose that we can describe an observable $\mathcal{O}$ as a function of the model.

- The expectation value for the observable can be written as

$$\mathsf{E}[\mathcal{O}] = \int d^n\alpha \mathcal{P}(\vec{\alpha})\mathcal{O}(\vec{\alpha}) = \frac{1}{N}\sum_k \mathcal{O}(\vec{\alpha}_k) \tag{13}$$

- and the variance is given by

$$\mathsf{Var}[\mathcal{O}] = \frac{1}{N}\sum_k (\mathcal{O}(\vec{\alpha}_k) - \mathsf{E}[\mathcal{O}])^2 \tag{14}$$

The recipe for reweighting

- With the new evidence D we can replace $\mathcal{P}(\vec{\alpha})$ by $\mathcal{P}(\vec{\alpha}|D)$.

$$\mathsf{E}[\mathcal{O}] = \int d^n\alpha \mathcal{P}(\vec{\alpha}|D)\mathcal{O}(\vec{\alpha})$$

$$= \int d^n\alpha \frac{\mathcal{P}(D|\vec{\alpha})}{\mathcal{P}(D)}\mathcal{P}(\vec{\alpha})\mathcal{O}(\vec{\alpha})$$

$$= \frac{1}{N}\sum_k w_k \mathcal{O}(\vec{\alpha}_k) \quad (15)$$

$$\mathsf{Var}[\mathcal{O}] = \frac{1}{N}\sum_k w_k (\mathcal{O}(\vec{\alpha}_k) - \mathsf{E}[\mathcal{O}])^2 \quad (16)$$

- Notice that $\mathcal{O}(\vec{\alpha_k})$ is sampled with the prior distribution.

- Method 1

$$\mathcal{P}(\vec{\alpha}|\vec{y}) = \frac{\mathcal{P}(\vec{y}|\vec{\alpha})}{\mathcal{P}(\vec{y})}\mathcal{P}(\vec{\alpha})$$

$$\Downarrow$$

$$w_k \;\propto\; \exp\left(-\frac{1}{2}\chi^2(\vec{\alpha_k}, \vec{t_k})\right)$$

- Method 2

$$\mathcal{P}(\vec{\alpha}|\chi) = \frac{\mathcal{P}(\chi|\vec{\alpha})}{\mathcal{P}(\chi)}\mathcal{P}(\vec{\alpha})$$

$$\Downarrow$$

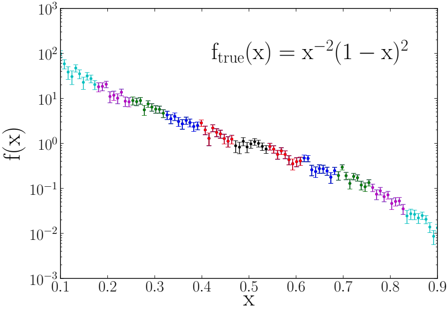
$$w_k \;\propto\; \exp\left(-\frac{1}{2}\chi^2(\vec{\alpha_k}, \vec{t_k})\right) \times \left(\chi^2(\vec{\alpha_k}, \vec{t_k})\right)^{\frac{1}{2}(n-1)}$$

- Q: Which method is better?

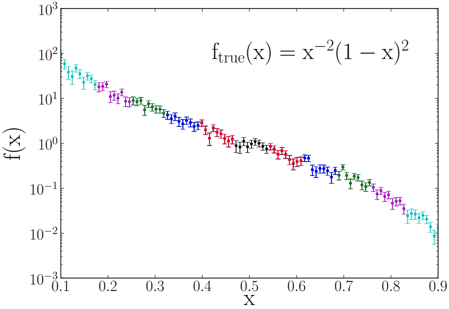
Simple numerical example:

1. Construct simulated data from $f(x,\vec{\alpha}) = x^{-2}(1-x)^2$ using Gaussian noise with uncorrelated errors.
2. Fit a model $f(x,\vec{\alpha}) = x^{\alpha_0}(1-x)^{\alpha_1}$ with a subset of the simulated data.
3. Get a Monte Carlo sample $\{\vec{\alpha}_k\}$.
4. Get predictions for a different subset of the simulated data for each $\vec{\alpha}_k$.

1. Compute the weights $\{w_k\}$.
2. Obtain expectation values and variances.
3. Compare the results with a fit that include both data sets.



Simple numerical example: simulated data



- For the analysis we split the data as follows

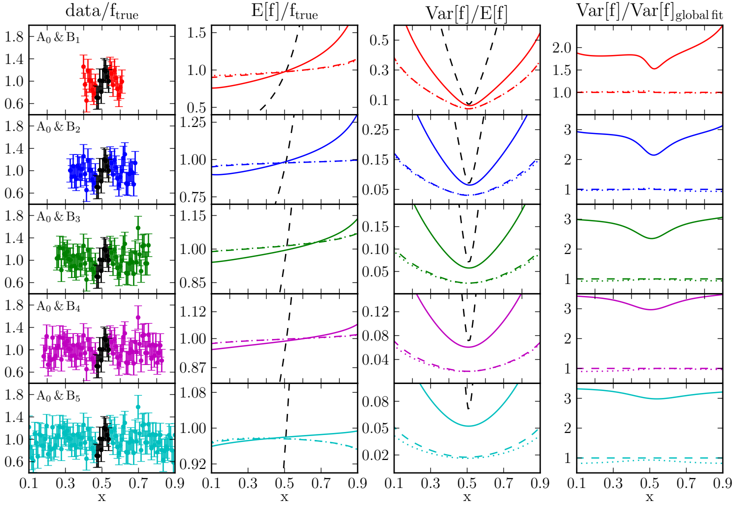
SET	data
A_0	d_5
A_1	d_4, d_5, d_6
A_2	d_3, d_4, d_5, d_6, d_7
A_3	$d_2, d_3, d_4, d_5, d_6, d_7, d_8$
A_4	$d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9$
A_5	$d_0, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}$

SET	data
B_1	d_4, d_6
B_2	d_3, d_4, d_6, d_7
B_3	$d_2, d_3, d_4, d_6, d_7, d_8$
B_4	$d_1, d_2, d_3, d_4, d_6, d_7, d_8, d_9$
B_5	$d_0, d_1, d_2, d_3, d_4, d_6, d_7, d_8, d_9, d_{10}$

SET	data
C_5	d_0, d_{10}

- $A_0 \equiv$ black color data.
- $A_1 \equiv$ red + black color data, etc.

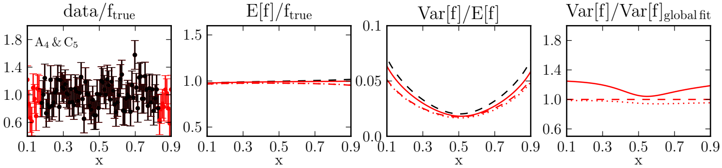
Simple numerical example: reweighting set A_0 with sets B_i



- Dashed: global fit (black is the fit with only A_0)
- Dotted: reweighting with method 1 (likelihood $\propto \mathcal{P}(\vec{y}|\vec{\alpha})$)
- Solid: reweighting with method 2 (likelihood $\propto \mathcal{P}(\chi|\vec{\alpha})$)

Simple numerical example: reweighting set A_4 with sets C_5

- Q: What if the initial data constrains well the parameters $\vec{\alpha}$?



- A: The two methods yield statistically equivalent results.

Conclusions:

- The method 1 is more efficient than method 2 as expected.
- Reweighting method 1 is statistically equivalent to global fit.
- The method 2 is equivalent to global fit in the limit where the prior parameters are well constrained.

The NNPDF paradox

- The NNPDF collaboration argues that the reweighting with method 1 is incorrect (see arxiv:1012.0836, arxiv:1108.1758).
- Consider $n-$dimensional space where n is the number of data points and the origin is at the prior predictions $\vec{t}(\vec{\alpha_0})$ for the new evidence.
- The distance for a given point $\vec{y}$ to the origin is given by $\chi^2 = (\vec{y} - \vec{t})^T \ \Sigma^{-1}(\vec{y} - \vec{t})$. Notice that Σ^{-1} is the metric for this space.
- Sets of constant χ^2 are $n-1$ dimensional surfaces. The NNPDF collaboration integrates the surfaces which gives method 2 for the reweighting.
- This is equivalent to ignoring the direction of the new evidence in the $n-$dimensional space, and therefore having less information.

The recipe for reweighting PDFs

The recipe for reweighting PDFs

- Typically, for hadron collisions, the cross sections are given by

$$\sigma(\tau) = \int_0^1 dx_a f_{a/A}(x_a) \int_0^1 dx_b f_{b/B}(x_b) \int_0^1 d\hat{\tau}\ \hat{\sigma}_{a,b}(\hat{\tau})\ \ \delta\left(\tau - \hat{\tau} x_a x_b\right)$$

- A random PDF can be written as

$$f^k(x) = f^0(x) + \frac{t}{2} \sum_j [f_j^+(x) - f_j^-(x)] R_j^k \tag{17}$$

- Then a random cross section will be given by

$$\sigma^k(\tau) = \int_0^1 dx_a f^k_{a/A}(x_a) \int_0^1 dx_b f^k_{b/B}(x_b) \int_0^1 d\hat{\tau}\ \hat{\sigma}_{a,b}(\hat{\tau})\ \ \delta\left(\tau - \hat{\tau} x_a x_b\right)$$

- Comparing with experimental cross sections we obtain $\{w_k\}$. We can finally obtain the reweighted PDFs:

$$\mathsf{E}[f_a] = \frac{1}{N} \sum_k w_k f_a^k \qquad \text{and} \qquad \mathsf{Var}[f_a] = \frac{1}{N} \sum_k w_k (f_a^k - \mathsf{E}[f_a])^2 \tag{18}$$

The recipe for reweighting PDFs

- The statistical convergence of the reweighting depends on the number of Monte Carlo samples.

- Instead of using

$$\sigma^k(\tau) = \int_0^1 dx_a f^k_{a/A}(x_a) \int_0^1 dx_b f^k_{b/B}(x_b) \int_0^1 d\hat{\tau}\ \hat{\sigma}_{a,b}(\hat{\tau})\ \ \delta\left(\tau - \hat{\tau} x_a x_b\right)$$

- we should use

$$\sigma^k(\tau) = \sigma_{00}(\tau) + \frac{t}{2}\sum_j \sigma_{0j}(\tau) R^k_j + \frac{t^2}{4}\sum_j \sigma_{ij}(\tau) R^k_i R^k_j \qquad (19)$$

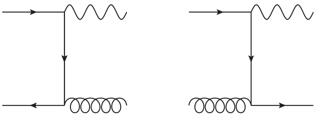
- σ_{00} is the calculation using the central PDFs
- σ_{0j} uses a central PDF and the difference in the j-th eigen PDFs
- σ_{ij} uses differences in the i-th and j-th eigen PDFs

Example:

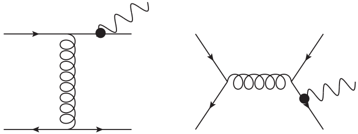
- data: single inclusive direct photon data from fixed target experiments.
- Due to inconsistencies between the data and its theory predictions @ NLO in pQCD, the data were excluded from global fits.
- Better theoretical description using threshold resummation @ NLO+NLL is available.
- prior PDFs: CJ12min
- for pp and pN collisions, direct photon cross sections are sensitive to initial state gluons.
- Currently, the information on the gluon PDF comes from Jet data. Yet its uncertainties are still large in the kinematic region of direct photon data.

Theory of direct photons

At LO:



(a) direct contribution



(b) jet fragmentation

$$p_T^3 \frac{d\sigma(x_T)}{dp_T} = \sum_{a,b,c} f_{a/A}(x_a, \mu_{IF}) * f_{b/B}(x_b, \mu_{IF}) * D_{\gamma/c}(z, \mu_{FF}) * \hat{\Sigma}(\hat{x}_T, ...)$$

- Direct contribution: $D_{\gamma/\gamma} = \delta(1-z)$
- Jet fragmentation: $D_{\gamma/c} \sim \alpha_{em}/\alpha_S$

Theory of direct photons

Beyond LO:

$$p_T^3 \frac{d\sigma(x_T)}{dp_T} = \sum_{a,b,c} f_{a/A}(x_a, \mu_{IF}) * f_{b/B}(x_b, \mu_{IF}) * D_{\gamma/c}(z, \mu_{FF}) * \hat{\Sigma}(\hat{x}_T, ...)$$

$\hat{\Sigma}(\hat{x}_T, ...) \supset$

1				LO
$\alpha_s L^2$	$\alpha_s L$	α_s		NLO
$\alpha_s^2 L^4$	$\alpha_s^2 L^3$	$\alpha_s^2 L^2$	$\alpha_s^2 L$	NNLO
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\alpha_s^n L^{2n}$	$\alpha_s^n L^{2n-1}$	$\alpha_s^n L^{2n-2}$	...	N^{n}LO
LL	NLL	NNLL	...	

$\hat{x}_T = 2p_T / z\sqrt{\hat{s}}$

$\hat{s} = x_a x_b S$

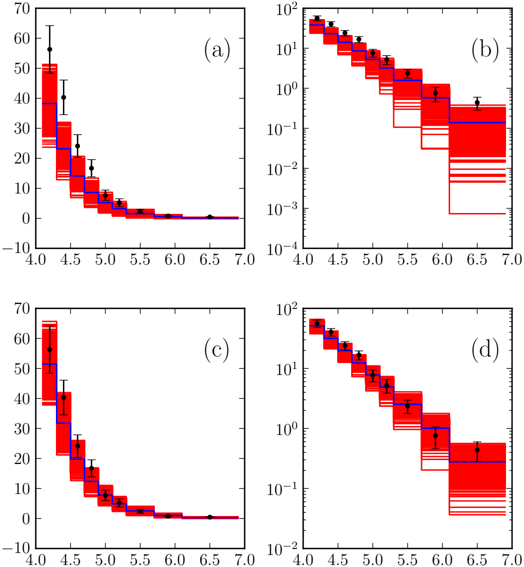
$L = \ln(1 - \hat{x}_T^2)$ "Threshold logs"

if $\alpha_S \; L^2 \sim 1$ resummation is needed.

Preliminary results:

reweighting using UA6 data

Example: single inclusive direct photon data UA6 pp



$\mathrm{E}\ \mathrm{d}\sigma/\mathrm{dp_T}\ (\mathrm{pb})$ vs $\mathrm{p_T}\ (\mathrm{GeV})$

(a) & (b) : NLO

(c) & (d) : NLO + NLL

$\chi^2_{\mathrm{DOF}}\,(\mathrm{NLO}) = 3.9$

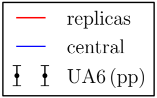
$\chi^2_{\mathrm{DOF}}\,(\mathrm{NLO+NLL}) = 0.8$

$\sqrt{\mathrm{s}} = 24.3$ GeV

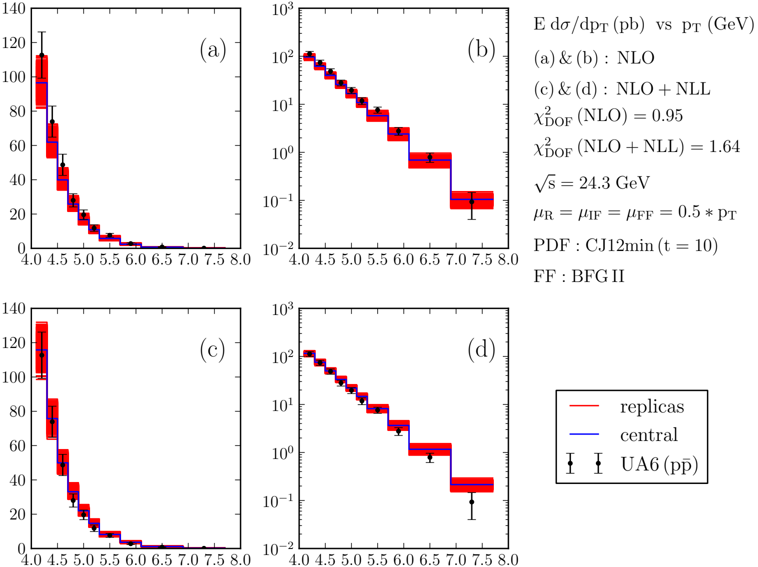
$\mu_\mathrm{R} = \mu_\mathrm{IF} = \mu_\mathrm{FF} = 0.5 * \mathrm{p_T}$

PDF : CJ12min (t = 10)

FF : BFG II



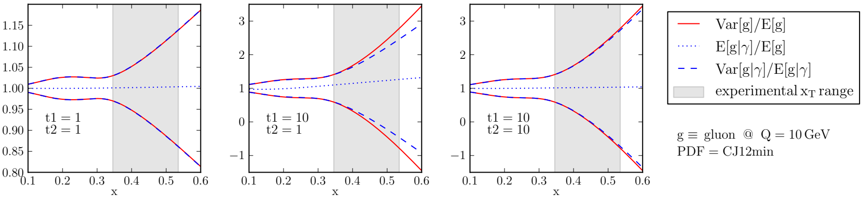
Example: single inclusive direct photon data UA6 $p\bar{p}$



Note about tolerance factor

- CTEQ, MSTW, uses a tolerance criterion.
- The idea is to define an acceptable region in the vicinity around minimum of the χ^2 such that $\Delta\chi^2 < t$.
- Then the uncertainties in the PDFs are enhanced by a factor of t.
- This procedure can be mimicked by the reweighting method in two ways:
  1. reweight PDFs with $t = 1$ and then enhance the uncertainty by a factor of t.
  2. replace the weights by $\chi^2_k \to \frac{1}{t}\chi^2_k$

Example: single inclusive direct photon data UA6 pp



- $t1$: a tolerance factor for the PDF uncertainty.
- $t2$: a factor to modify the weights.
- The normalization uncertainty is included using the χ^2

$$\chi^2 = \sum_i \left(\frac{D_i + nD_i\lambda - T_i}{\sigma_i} \right)^2 + \lambda^2. \tag{20}$$

- $n = 11\%$ for UA6 data
- Full analysis including more data sets (WA70,E706,ISR,...) is in preparation.

Conclusions

- A reweighting technique with 2 different prescriptions has been presented.
- A simple numerical exercise, shows that one of the methods is statistically equivalent with global fits.
- A recipe to reweight non Monte Carlo based PDFs such as CTEQ, MSTW has been presented.
- Some preliminary results on PDF reweighting using single inclusive direct photon data have been shown.