New Generation of Parton Distribution Functions with Uncertainty Analysis

Dan Stump
Michigan State University

CTEQ6

Two advances

• treatment of experimental systematic errors

• a full uncertainty analysis


hep-ph/0201195
Related Work


Parton Distribution Functions

\[ \sigma(Q^2) = \sum_i \int_0^1 \hat{\sigma}_i(Q^2) f_i(x, Q^2) \, dx \]

Global analysis

- Parametrize \( f_i(x, Q_0^2; a_1 \ldots a_n) \) shape parameters
- NLO evolution in \( Q \implies f_i(x, Q^2) \)
- The NLO parton cross sections (\( \hat{\sigma}_i \)) are known.
- Find the “optimum” parameter values to fit data for many short-distance processes.
## Selection of Data

<table>
<thead>
<tr>
<th></th>
<th>CTEQ5</th>
<th></th>
<th>CTEQ6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#</td>
<td>sys</td>
<td>#</td>
<td>sys</td>
</tr>
<tr>
<td>BCDMS $\mu p$</td>
<td>168</td>
<td>no</td>
<td>BCDMS $\mu p$</td>
<td>339</td>
</tr>
<tr>
<td>BCDMS $\mu d$</td>
<td>156</td>
<td>no</td>
<td>BCDMS $\mu d$</td>
<td>251</td>
</tr>
<tr>
<td>H1 $e p$</td>
<td>172</td>
<td>no</td>
<td>H1a $e p$</td>
<td>104</td>
</tr>
<tr>
<td>ZEUS $e p$</td>
<td>186</td>
<td>no</td>
<td>ZEUS $e p$</td>
<td>229</td>
</tr>
<tr>
<td>NMC $\mu p$</td>
<td>104</td>
<td>no</td>
<td>NMC $\mu p$</td>
<td>201</td>
</tr>
<tr>
<td>NMC $\mu p/\mu n$</td>
<td>123</td>
<td>no</td>
<td>NMC $\mu p/\mu n$</td>
<td>123</td>
</tr>
<tr>
<td>CCFR $F_2 \nu N$</td>
<td>87</td>
<td>no</td>
<td>CCFR $F_2 \nu N$</td>
<td>159</td>
</tr>
<tr>
<td>CCFR $F_3 \nu N$</td>
<td>87</td>
<td>no</td>
<td>CCFR $F_3 \nu N$</td>
<td>87</td>
</tr>
<tr>
<td>E605 $p p$ DY</td>
<td>119</td>
<td>no</td>
<td>E605 $p p$</td>
<td>119</td>
</tr>
<tr>
<td>NA51 $p d/p p$ DY</td>
<td>1</td>
<td>no</td>
<td>NA51 $p d/p p$</td>
<td>1</td>
</tr>
<tr>
<td>E866 $p d/p p$ DY</td>
<td>15</td>
<td>no</td>
<td>E866 $p d/p p$</td>
<td>15</td>
</tr>
<tr>
<td>CDF W</td>
<td>11</td>
<td>no</td>
<td>CDF W</td>
<td>11</td>
</tr>
<tr>
<td>CDF jet</td>
<td>33</td>
<td>yes</td>
<td>CDF jet</td>
<td>33</td>
</tr>
<tr>
<td>DØJet</td>
<td>24</td>
<td>yes</td>
<td>DØJet</td>
<td>90</td>
</tr>
</tbody>
</table>
\( \chi^2 \) and Systematic Errors

The simplest definition

\[
\chi^2_0 = \sum_{i=1}^{N} \frac{(D_i - T_i)^2}{\sigma_i^2}
\]

\[
\begin{cases}
D_i = \text{data} \\
T_i = \text{theory} \\
\sigma_i = \text{"expt. error"}
\end{cases}
\]

is optimal for random Gaussian errors,

\[
D_i = T_i + \sigma_i r_i \quad \text{with} \quad P(r) = \frac{e^{-r^2/2}}{\sqrt{2\pi}}.
\]

With systematic errors,

\[
D_i = T_i(a) + \alpha_i r_{\text{stat},i} + \sum_{k=1}^{K} r_k \beta_{ki}.
\]

The fitting parameters will be \( \{a_\lambda\} \) (theoretical model) and \( \{r_k\} \) (corrections for systematic errors).

Published experimental errors:

- \( \alpha_i \) is the ‘standard deviation’ of the random uncorrelated error.

- \( \beta_{ki} \) is the ‘standard deviation’ of the \( k \) th (completely correlated!) systematic error on \( D_i \).
To take into account the systematic errors, we define

$$\chi'^2(a_\lambda, r_k) = \sum_{i=1}^{N} \frac{(D_i - \sum_k r_k \beta_{ki} - T_i)^2}{\alpha_i^2} + \sum_k r_k^2,$$

and minimize with respect to \(\{r_k\}\). The result is

$$\hat{r}_k = \sum_{k'} (A^{-1})_{kk'} B_{k'}, \quad \text{(systematic shift)}$$

where

$$A_{kk'} = \delta_{kk'} + \sum_{i=1}^{N} \frac{\beta_{ki} \beta_{k'i}}{\alpha_i^2}$$

$$B_k = \sum_{i=1}^{N} \frac{\beta_{ki} (D_i - T_i)}{\alpha_i^2}.$$

Note that the \(\hat{r}_k\)'s depend on the PDF model parameters \(\{a_\lambda\}\). Then

$$\chi^2(a_\lambda) = \min_{\{r_k\}} \chi'^2(a_\lambda, r_k)$$
Now minimize $\chi^2(a)$ with respect to the model parameters $\{a_\lambda\}$.

**Output**

- $\{a_\lambda\}$, which determine $f_i(x, Q^2_0)$.
- $\{\hat{r}_k\}$, which are optimal “corrections” for systematic errors; i.e., systematic shifts to be applied to the data points to bring the data from different experiments into compatibility, within the framework of the theoretical model.
Overview of the CTEQ6M parton distribution functions at $Q = 2$ and 100 GeV.
Comparison of CTEQ6M (dashed) to CTEQ5M1 (dot-dashed) PDF's at $Q = 2$ GeV. (The unlabeled curves are $\bar{u}$ and $s = \bar{s}$.)

★ Quarks have not changed much.

★ Gluon is noticeably different.
Comparison of CTEQ6M (dashed) to CTEQ5M1 (dot-dashed) gluon distributions at $Q = 2$ and 100 GeV. (a) The small-$x$ region; (b) the large-$x$ region.
Comparison to Data

Comparison of the CTEQ6M fit to data with correlated systematic errors.

<table>
<thead>
<tr>
<th>data set</th>
<th>$N_e$</th>
<th>$\chi_e^2$</th>
<th>$\chi_e^2/N_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCDMS p</td>
<td>339</td>
<td>377.6</td>
<td>1.114</td>
</tr>
<tr>
<td>BCDMS d</td>
<td>251</td>
<td>279.7</td>
<td>1.114</td>
</tr>
<tr>
<td>H1a</td>
<td>104</td>
<td>98.59</td>
<td>0.948</td>
</tr>
<tr>
<td>H1b</td>
<td>126</td>
<td>129.1</td>
<td>1.024</td>
</tr>
<tr>
<td>ZEUS</td>
<td>229</td>
<td>262.6</td>
<td>1.147</td>
</tr>
<tr>
<td>NMC F2p</td>
<td>201</td>
<td>304.9</td>
<td>1.517</td>
</tr>
<tr>
<td>NMC F2d/p</td>
<td>123</td>
<td>111.8</td>
<td>0.909</td>
</tr>
<tr>
<td>DØ jet</td>
<td>90</td>
<td>64.86</td>
<td>0.721</td>
</tr>
<tr>
<td>CDF jet</td>
<td>33</td>
<td>48.57</td>
<td>1.472</td>
</tr>
</tbody>
</table>

Other data sets:

- CCFR $\nu$ DIS (150/156)
- E605 Drell-Yan (95/119)
- E866 Drell-Yan (6/15)
- CDF W-lepton asymmetry (10/11)
Comparison of the CTEQ6M fit to the ZEUS data in separate $x$ bins.* The data points include the estimated corrections for systematic errors. The error bars are statistical errors only.

Histogram of residuals for the ZEUS data. The curve is a Gaussian of width 1. (a) $\Delta_i = (D_i - \sum_k \hat{r}_k \beta_{ki} - T_i)/\alpha_i$. (b) A similar comparison but without the corrections for systematic errors on the data points.

$$D_i^{\text{corrected}} = D_i - \sum_{k=1}^{K} \hat{r}_k \beta_{ki}$$
Systematic shifts for the ZEUS data (10 systematic errors)

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{r}_k$</td>
<td>1.67</td>
<td>-0.67</td>
<td>-1.25</td>
<td>-0.44</td>
<td>-0.00</td>
<td>-1.07</td>
<td>1.28</td>
<td>0.62</td>
<td>-0.40</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Systematic shifts for the NMC data (11 systematic errors)

<table>
<thead>
<tr>
<th>$k$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{r}_k$</td>
<td>0.67</td>
<td>-0.81</td>
<td>-0.35</td>
<td>0.25</td>
<td>0.05</td>
<td>0.70</td>
<td>-0.31</td>
<td>1.05</td>
<td>0.61</td>
<td>0.26</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Comparison of the CTEQ6M fit with the BCDMS and NMC data on $\mu p$ DIS.\,*

Histogram of residuals for the NMC data.

(a) $\Delta_i = \frac{(D_i - \sum_{k=1}^{K} \hat{\tau}_{k,\beta_{ki}} - T_i) / \alpha_i}{\omega_i}$.

(b) A similar comparison but without the corrections for systematic errors on the data points.

$D_{\text{Corrected}} = D_i - \sum_{k=1}^{K} \hat{\tau}_{k,\beta_{ki}}$.
Comparison of the CTEQ6M fit to the inclusive jet data.* (a) DØ (the boundary values of the 5 rapidity bins are 0, 0.5, 1.0, 1.5, 2.0 and 3.0). (b) CDF (central rapidity, 0.1 < |η| < 0.7).

Closer comparison between CTEQ6M and the DØ jet data as fractional differences.

The Tevatron inclusive jet cross section implies a hard gluon: $g(x)$ is large at large $x$.
Recall CTEQ4HJ and CTEQ5HJ.
CDF inclusive jet cross section

Quantitative Uncertainties

A SIMPLE IDEA

\[ \chi^2 \text{ global} \]

\[ \chi^2_0 + \Delta \chi^2 \]

\[ \chi^2_0 \]

\[ X_0 - \Delta X \quad X_0 \quad X_0 + \Delta X \]

**Tolerance**

\[ T^2 = \max \text{ allowed } |\Delta \chi^2| \]

\[ T \approx 10 \]

**COMPUTATIONAL METHODS**

- Lagrange Multiplier Method – constrained minimization to obtain the best fit as a function of \( X \).
- Hessian Matrix Method – explore the variation of \( \chi^2(a) \) in the neighborhood of the global minimum in the \( n \) dimensional parameter space.

**2-dim (i,j) rendition of d-dim (~20) PDF parameter space**

- \( u_l \): eigenvector in the \( l \)-direction
- \( p(i) \): point of largest \( a_i \) with tolerance \( T \)
- \( s_0 \): global minimum

**Original parameter basis**

**Orthonormal eigenvector basis**

**Diagonalization and rescaling by the iterative method**

**Hessian eigenvector basis sets**
Uncertainty bands for the $u$- and $d$-quark distribution functions at $Q^2 = 10 \text{ GeV}^2$. The solid line is CTEQ5M1 and the dotted line is MRST2001.

(Shaded: *envelopes* of extreme pdf's)
Uncertainty band for the gluon distribution function at $Q^2 = 10 \text{ GeV}^2$. The curves correspond to CTEQ5M1(solid), CTEQ5HJ (dashed), and MRST2001 (dotted).

The gluon is very uncertain for $x \gtrsim 0.4$.

$[g(x) \to 0 \text{ as } x \to 1.]$
Uncertainties of the luminosity functions at the Tevatron.

\[ \mathcal{L}(\hat{s}) = \sum_{i,j} C_{ij} \int f_i(x_1) f_j(x_2) \delta(\hat{s} - x_1 x_2 s) dx_1 dx_2 \]
Uncertainties of the luminosity functions at the LHC.
A knotty problem

What's the tolerance?

Our analysis: Consider the $\chi^2$'s of individual experiments, as any variable changes from its value at the global minimum.

![Eigenvector 4 diagram](image)

Uncertainty ranges (vertical lines) for the input experiments along the Eigenvector 4 direction. The horizontal lines indicate the tolerance $T$. (Points: positions of minimum $\chi^2$ for individual experiments)

(distance $= \sqrt{\Delta \chi^2_{\text{global}}}$)
$\Delta \chi^2$ of individual data sets versus distance along Eigenvector 4.
The red line is the 90% confidence range for $\chi^2_{s\text{caled}}$.

$$\left( \chi^2_{s\text{caled}} \equiv \frac{\chi^2}{\chi_0^2} \right)$$
Uncertainty ranges for the input experiments along the Eigenvector 18 direction.

A *reasonable* (as opposed to strict) tolerance criterion leading to $T \approx 10$:

- take the *intersection* of the individual 90% confidence ranges . . .
- . . . but a single anomalous experiment cannot set the bound.

The uncertainties on predictions are $\propto T$. 