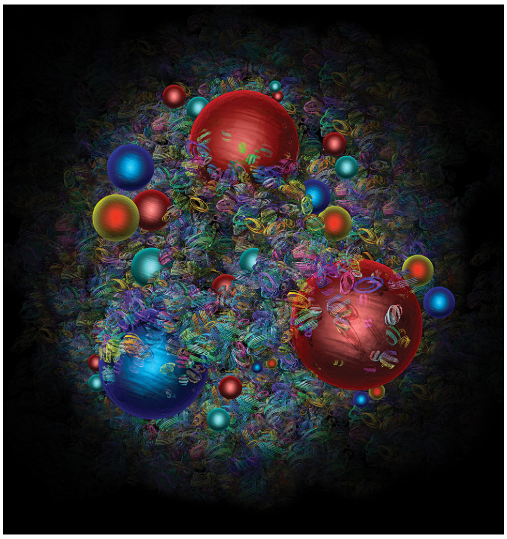
DEPARTMENT OF Physics
Florida State University,
Tallahassee, Florida

PDFs Parametric Uncertainty Studies



A Thesis

Submitted by

ALI AL KADHIM

For the Fullfilment

Of

SUMMER 2021 RESEARCH PROJECT

July 2021

QUOTATIONS

In science, progress is possible.
In fact, if one believes in Bayes'
theorem, scientific progress is
inevitable as predictions are made
and as beliefs are tested and
refined.

NATE SILVER

ACKNOWLEDGEMENTS

Thanks to Dr. Prosper for everything.

ABSTRACT

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i

Page

ABSTRACT ii

LIST OF TABLES vii

LIST OF FIGURES viii

GLOSSARY ix

ABBREVIATIONS x

NOTATION xi

CHAPTER 1: Introduction and Motivation 1

CHAPTER 2: PDFs and QCD: Introduction, theory, determination . . . 2

2.1 What are Parton Distribution Functions? 2

2.1.1 Relevant Fundamentals of QCD 2

2.1.2 Introduction on PDFs 3

2.1.3 Peculiarities about PDFs 6

2.1.4 Deep Ineslastic Scaterring 6

2.1.5 The Usual Determination of PDFs 7

CHAPTER 3: Relevant Statistics 10

3.1 General Methods 10

3.1.1 Probability Distributions 10

3.1.2 Method of Least Squares 11

3.1.3 Maximum Likelihood Method 12

3.1.4 Method of Maximum Likelihood (Cowan) 13

3.1.5 Chi-squared Test 14

3.1.6 Confidence Intervals 16

3.1.7 Multivariate and χ^2 18
3.2 Bayesian Reweighting 20
3.2.1 Geile and Keller (GK) Method 22
3.2.2 chi-squared 24
3.3 The Hessian method 25
3.3.1 The Hessian Method written another way 30
3.4 The Bootstrap 32
3.5 Closure Test 32
3.6 To Do 33
3.7 Notes on Papers 33
3.7.1 Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies 33
3.7.2 A First Determination of Parton Distributions with Theoretical Uncertainties 33
3.7.3 Parton distributions for the LHC 33
3.7.4 Parton distributions for the LHC Run II 34
3.7.5 xFitter 2.0.0: An Open Source QCD Fit Framework 34
3.7.6 Uncertainties of predictions from parton distribution functions II: the Hessian method 34
3.7.7 Closure Testing the NNPDF3.0 methodology 34
3.7.8 Bayesian Reweighting for Global Fits - Prosper 35
3.7.9 Updating and Optimizing Error PDFs in the Hessian Approach 35
3.7.10 A new simple PDF parametrization: improved description of the HERA data 36
3.7.11 PDF reweighting in the Hessian matrix approach - 2014 . . 36
3.7.12 Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO 37
CHAPTER 4: Problems in PDFs: uncertainties, extrapolation, MHOTs,. . .
38
4.1 Problems of Uncertainty 38
4.1.1 Missing Higher Order Uncertainty 39

4.1.2 Fitting Procedure in xFitter 40

4.1.3 Parameterization: How different Parameterizations yield different PDFs 41

4.1.4 Comparisons of using different experimental data 42

4.1.5 Using xFitter 42

4.1.6 Fast Interpolation Grids 43

4.1.7 xFitter Tutorials 43

4.2 xFitter 44

4.2.1 Fitting Procedure in xFitter 44

4.2.2 Parameterization: How different Parameterizations yield different PDFs 45

4.2.3 Comparisons of using different experimental data 47

4.2.4 Using xFitter 47

4.2.5 Fast Interpolation Grids 47

4.2.6 xFitter Tutorials 47

CHAPTER 5: NNPDF Approach 49

5.1 PDF Determination as a Pattern Recognition Problem 49

5.1.1 Differences From Standard ML applications 50

5.1.2 NNPDF The Software 51

5.1.3 NN architecture 51

5.1.4 Closure Testing by the NNPDF Collaboration 51

5.1.5 Future of NNPDFs: more hyperoptimization and code-redesign, apparently 52

CHAPTER 6: Testing PDF uncertainties within xFitter and closure Testing (our approach) 53

6.1 Statement of the Problem 53

6.2 Are the PDF parameters normally-distributed? 54

6.3 Determination of $\Delta\chi^2$ 56

6.4 Results 56

APPENDIX A: APPENDIX 57

A.1 More statistics 57

A.2 Hypothesis Tests 57

APPENDIX B: ANOTHER SAMPLE APPENDIX 59

REFERENCES 60

LIST OF TABLES

Table	Title	Page

LIST OF FIGURES

Figure Title Page

2.1 $p_A\,p_A$ collision figure 4

3.1 Caption 21
3.2 Visualizing the Hessian method: the Hessian matrix is diagonalized and its eigenvalues are rescaled 29

6.1 Caption 55

GLOSSARY

The following are some of the commonly used terms in this thesis:

OpenFOAM An opensource C++ toolbox for the development of customized numerical solvers, and pre-/post-processing utilities for the solution of continuum mechanics problems, most prominently including computational fluid dynamics

CFD A branch of fluid mechanics that uses numerical analysis and data structures to analyze and solve problems that involve fluid flows

FireFOAM FireFOAM is a CFD solver used for LES modeling of fire and its suppression in the OpenFOAM

ABBREVIATIONS

Florida State University

NOTATION

Greek Symbols

Miscellaneous

CHAPTER 1

Introduction and Motivation

CHAPTER 2

PDFs and QCD: Introduction, theory, determination

2.1 What are Parton Distribution Functions?

2.1.1 Relevant Fundamentals of QCD

The QCD Lagrangian is

$$\mathcal{L} - -\frac{1}{4} F^{A}_{\mu v} F^{\mu v}_{A} + \sum_{\text{flavous}} \bar{\psi}_a \left(i\gamma_\rho D^\mu - m \right)_{ab} \psi_b \tag{2.1}$$

where $F^{A}_{\mu\nu}$ is the gluon field strength

$$F^{A}_{\mu\nu} = \partial_\mu A^{A}_{\nu} - \partial_\nu A^{A}_{\mu} + g_s f^{ABC} A^{B}_{\mu} A^{C}_{\nu} \tag{2.2}$$

where D_μ is the covariant derivative

$$\left(D_\mu \right)_{ab} = \partial_\mu \delta_{ab} - i g_s A^{A}_{\mu} t^{A}_{ab} \tag{2.3}$$

Where t^A are the algebra generators. Two very important of QCD are confinement and Asymptotic Freedom. Confinement is the propery that no isolated coloured charge (like a quark) can exist as a free particle but only colour singlet particles can be isolated. For example, a proton is made up of 4 quarks (2 u's and 1 d) in a neutral colour configuration. All observed hadrons (particles with strong interaction) are either made up of 3 quarks (baryoons) or a quark-antiquark pair (mesons), all singlets under colour. Confinement is due to the fact that the potential between two coulour charges, for example a quark and antiquark, has a coulomb-like part at short distances, but a linearly rising term at long distances. The linearly rising term makes it energetically impossible to separate two coloured charges in the bound state. At some point in trying to separate them, it will become energetically favorable to create a quark-antiquark pair at the two end, just like when stretching a line made of a rubber band, at some point the rubber band will be cut into two smaller pieces.

Asymptotic freedom is the property that the QCD coupling $\alpha_s = g_s^2/4\pi$ becomes weak at high energies, due to quantum corrections, so that the theory becomes perturbative in this regime (that is, the theoretical predictions can be expressed as an expansion in powers of the coupling limited to the first few terms). Hence QCD has a low energy regime, in which the theory is strongly-niteracting and a high-energy one, in which it is asymptotically free. At energies large enough that masses can be neglected, naively one would expect that dimensionless measurable quantities would become “scale invariant”, namely independent of the absolute scale of energy, and only functions of energy ratios (“scaling variables”).

A "hard" process is one that occurs at high energies, and for it all energy variables are large and of the same scale; we denote the common energy scale by Q. Hard processes are also "infrared safe", that is, it is well defined in the limit of vanishing quark masses, and free of infrared singularities (that rise with the masses of the quarks approaching zero). For hard processes, any measureable quantity can only depend on Q and on a number of scaling variables x_i (ie we have "Bjorken scaling"). In reality, scaling is broken by QCD quantum corrections and renormalization requires a scale of mass Λ_{QCD}, but the scaling violations are only logarithmic and computable.

2.1.2 Introduction on PDFs

Parton distribution functions, which characterize the structure of the proton, are one of the most important sources of uncertainty in predictions of most observables in the LHC. The parton density function $f_i\left(x, Q^2\right)$ gives the probability of finding a parton (quark or gluon) of flavour i in the hadron (like a proton), where i is the different flavors (like up, down, etc.) carrying a fraction x of the proton’s momentum (so a parton will have momentum $p_{parton} = \hat{p} = xp$, where p is the momentum of the proton), with Q being the energy scale of the hard interaction (that the energy scale chosen). Note that the momentum fraction x is completely fixed by kinematics, i.e. by $Q^2 = -q^2$ and $W^2 = (p+q)^2$ where $q = k - k'$ (i.e. by $cos\theta$ and W^2). QCD does not predict the parton content of the proton, so the shapes of the PDFs are determined from experimental observable, where the cross sections are calculated by convoluting the parton level cross sections with the PDFs. The knowledge od proton PDFs mainly comes from Deep Inelastic Scattering from various particle collider experiments. Consider a hadron-hadron interaction where we have hadron A (with momentum P_A) collision with hadron B (with momentum p_B) leading to a jet and "unknown" X, ie we have the process $A\ B\ \rightarrow$ jet + X, as illustrated in Fig 2.1 The jet that we observe in the detector begins as a single quark or gluon that emerges from the parton-parton scattering event with a large p_T. The picture suggests that we could write the cross section of the produced jet as a product of three factors. These factors are two parton distribution functions $f_i(x_i)$ and a cross section

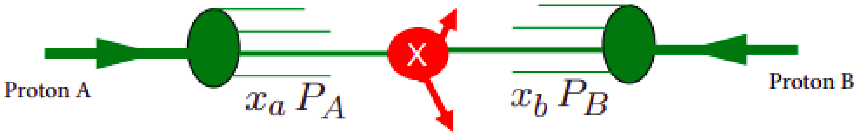


Fig. 2.1: $p_A\ p_A$ collision figure

- A parton of type a that comes from the hadron A. It carries a fraction x_A of the momentum of hadron A. The probability to find between momenta x_A and $x_A + dx_A$ it is given by the PDF $f_{a/A}(x_A)dx_A$.

- A second parton of type b that comes from the hadron B. It carries a fraction x_B of the momentum of hadron B. The probability to find between momenta x_B and $x_B + dx_B$ it is given by the PDF $f_{b/B}(x_B)dx_B$.

- The third factor is the cross section for the partons to make the observed jet, $d\sigma_{partons}$. This parton level cross section is calculated using perturbative QCD

Hence, in the language of QCD, the short-distance (high energy) part of the process can be computed from perturbation theory, and long-distance (low energy) part of the process is driven by the non-perturbative nature of QCD at low-energy scales. Collinear factorization theorem allows us to separate the perturbative (calculable) hard part of the process from the non-pertubative one, which can be described in terms of parton distribution (or fragmentation) functions. The total cross section of inelastic proton-

proton scattering to produce a final state n can be calculated with the formula

$$\begin{aligned}\sigma =& \sum_{a,b} \underbrace{\int_0^1 dx_a dx_b f_{a/A}\left(x_a, \mu_F\right) f_{b/B}\left(x_b, \mu_F\right)}_{\text{long-distance, non-perturbative PDF part}} \\ &\times \underbrace{\int d\Phi_n \frac{1}{2\hat{s}} \left|\mathcal{M}_{ab\to n}\right|^2 \left(\Phi_n; \mu_F, \mu_R\right)}_{\text{short-distance "hard" perturbative part}}\end{aligned} \tag{2.4}$$

Where $f_{a/A}(x, \mu)$ denotes the parton distribution functions, which depend on the momentum fractin x of a parton a with respect to its parent hadron A, and on an arbitrary energy scale called the factorization scale μ_F. $d\Phi_n$ is the differential phase space element over n final-state particles,

$$d\Phi_n = \prod_{i=1}^{n} \frac{d^3 p_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta^{(4)} \left(p_a + p_b - \sum_{i=1}^{n} p_i\right) \tag{2.5}$$

Where p_a and p_b are the initial state momenta. The convolution of the squared matrix element $\left|\mathcal{M}_{ab\to n}\right|^2$, averaged over initial-state spin and colour degrees of freedom, with the Lorentz-invariant phase space n and multiplied by the flux factor $1/(2\hat{}s) = 1/(2xaxbs)$ results in the calculation of the parton-level cross section $\hat{\sigma}_{ab ßn}$.

Hence we can intuitively say that the differential cross section in transverse momenta of the observed jet can be factorized in the following form [1]

$$\frac{d\sigma_{jet}}{dP_T} \sim \sum_{a,b} \int dx_a f_{a/A}\left(x_A, \mu\right) \int dx_b f_{b/B}\left(x_B, \mu\right) \frac{d\sigma_{partons}}{dP_T} \tag{2.6}$$

Where $\sigma_{partons} = \int d\Phi_n \frac{1}{2\hat{s}} \left|\mathcal{M}_{ab\to n}\right|^2 \left(\Phi_n; \mu_F, \mu_R\right)$ can be seen from 2.4. The equation 2.6 illustrates the principle of *factorization* [2]: i.e. that short distance and long distance processes are separable such that they can be convoluted in this manner, so that the "hard part" $\sigma_{partons}$ and "normalizations" from the PDFs are on diffrent scales. Factorization also posits that the PDFs are universal, i.e. process-independent. Factorization also states that since the strong coupling is running $\alpha_s = \alpha_s(\mu)$, equation 2.6 holds up to corrections of order

[1] sometimes this is called the "master formula"

[2] Beware that this equation is not strictly proven, it is proven hoewever for Drell-Yan processes (where dileptons are produced $pp \to l^+ l^-$

- $(m/p_T)^n$ where m is a typical hadronic mass scale and the power n depends on the process, and
- $(\alpha_s(\mu))^L$ from truncating the expansion of $d\sigma_{partons}/dp_T$

Note that the parameter μ (technically, the factorization or renormalization scale μ_F), which has dimensions of mass, is related to the renormalization of the strong coupling $\alpha_s(\mu)$ and to the operators in the definition of the parton distribution functions $f_{a/A}(x_A, \mu)$

The parton level cross sections $d\sigma_{partons}$ has an expansion in powers of α_S

$$\frac{d\sigma_{partons}}{dP_T} \sim \sum_N \left(\frac{\alpha_s(\mu)}{\pi}\right)^N H_N\left(x_A, x_B, P_T; a, b; \mu\right) \tag{2.7}$$

Where the coefficients H_N are calculable in perturbative QCD. Equation **??** demonstrates the principle of *Asymptotic Freedom*, i.e. hard scattering is weak at short di3 stances, and hence perturbatively calculable. At next-to-leading-order and beyond, however, the calculation will involve divergences that must be removed, and the dependence on the scale μ will appear in their place. Hence the picture is that measuring total cross section $\Longleftrightarrow$ need to know the PDFs to be able to test the hard part (for example the Higgs electroweak coupling).

2.1.3 Peculiarities about PDFs

PDFs are not quite probability densities, because they are not functions but rather distributions.

2.1.4 Deep Ineslastic Scaterring

The hadron-hadron proces is complex to describe, so let's start with a simpler case of Deep inelastic scattering (DIS), where electrons are collided with hadrons (usually protons). Two different deep inelastic ep scattering processes were measured by HERA: Neutral current (NC) $ep \to eX$, and charged current (CC) $ep \to \nu X$. In neutral current reactions the interaction proceeds with the exchange of a photon or a Z boson, whereas in charged current scattering a $W^{\pm}$ boson is exchanged. The DIS cross section depends on the structure functions, and on the proton and lepton helicities. For example, the cross section of ep can be written in the form

$$\frac{d^2\sigma^{\lambda_p\lambda_\ell}\left(x,y,Q^2\right)}{dxdy} = \frac{G_F^2}{2\pi\left(1+Q^2/m_W^2\right)^2}\frac{Q^2}{xy}\left\{\left[-\lambda_\ell y\left(1-\frac{y}{2}\right)xF_3\left(x,Q^2\right)+(1-y)F_2\left(x,Q^2\right)\right.\right.$$
$$\left.\left.+y^2xF_1\left(x,Q^2\right)\right]-2\lambda_p\left[-\lambda_\ell y(2-y)xg_1\left(x,Q^2\right)-(1-y)g_4\left(x,Q^2\right)-y^2xg_5\left(x,Q^2\right)\right]\right\} \tag{2.8}$$

Where λ_l is the lepton helicity and λ_p is the proton helicity.

In the quark parton model, the DIS cross section to zeroth order in α_s is

$$\frac{d^2\sigma^{em}}{dxdQ^2} \simeq \frac{4\pi\alpha^2}{xQ^4}\left(\frac{1+(1-y)^2}{2}F_2^{em}+\mathcal{O}\left(\alpha_\mathrm{s}\right)\right) \tag{2.9}$$

Where F_2 is a structure function.

$$F_2 = x\left(e_u^2u(x)+e_d^2d(x)\right) = x\left(\frac{4}{9}u(x)+\frac{1}{9}d(x)\right) \tag{2.10}$$

Where $u(x)$ and $d(x)$ are parton distribution functions.

2.1.5 The Usual Determination of PDFs

The PDFs could be computable if one was able to solve QCD in the non-perturbative domain, ie. if it was possible to compute the proton wave function from first principles. This is not the case (currently), and hence, PDFs are a set of well-defined functions of x at some reference scale Q_0, which depend on the only free parameter of the theory, α_s. In other words, we know that these functions exist, but we don't know what they are.

[3]

The general procedure that has been carried out so far to determine the PDFs is the following. Starting from a parameterisation of the non-perturbative PDFs at a low scale μ, either by making assumptions on their analytical form or by using the neural-net technology, fits to various sets of data (mainly to DIS data) are performed within the

[3] At present, the only way to determine them is by comparing criss sections in the form of 2.6 for a wide enough set of observables for which the hadronic cross section is measured with sufficient precision, and the partonic cross section is known with sufficient accuracy. **Question: since we have to have such accurate experimental measurements of finanl state cross sections, why are cross sectional measurements from old measurements ususlly used as opposed to new more accurate cross sectional measurements?**

DGLAP [4] evolution scheme. Hence a particular functional form of the x dependence of the PDFs at a given reference scale μ, by solving the perturbative evolution equations and determining the free parameters by fitting to the data. The standard choice of this functional form, which is suggested from theory arguments, is

$$f_{a/A} = x^{\alpha_{a/A}}(1-x)^{\beta_{a/A}} \tag{2.11}$$

There are some problems with this functional form

- The power-like behavior (as $x \to 0$ and $x \to 1$ is spurious; and even if it's true, there is no reason to believe that this will hold for all x. Hence given that only a finite range of x is experimentally accessible, there is no reason that this form applies in the experimentally-observable region.

- Even if the PDF takes the form of 2.11 at some scale, this form is not preserved as the scale is varied: specifically, it is corrected by lnx terms is $x \to 0$ and by $ln(1-x)$ terms as $x \to 1$.

- Uncertainties: uncertainties on the fit parameters determined by least-squares and standard error propagation turned out to be smaller by about one order of magnitude than one might reasonably expect by looking at the fluctuation of best-fit values as the underlying dataset was varied. Another problem associated with uncertainties is that the error bands with newer PDF sets were larger than old PDF sets, despite the newer PDF sets' parameterization being more complex (more parameters) and the data collected being more. This leads to the unusual and counter-intuitive effect in which adding new data seems to make uncertainty go up as opposed to down, counter to the situation in any other circumstance. As the PDF form with more parameters leads to higher uncertainty, this leads to the conclusion that uncertainties were underestimated by the bias [5] in the choice of parameterization.

Hence this simple form is quite restrictive and biased; and we can complicate it by forming more elaborate forms, but we are blind as to their exact analytical form. There are other problems with the old way of PDF determination as well:

- $x = 1$ is a kinematic boundary. But data stops at some $x < 1$, so we must interpolate PDFs to large x.

- The scale on x is naturally logarithmic, so it should go to $-\infty$, but data stops at some small x. It is unknown what goes beyond that point, hence we must extrapolate to low x.

[4] DGLAP stands for the physicists that came up with this scheme: Dokshitzer, Gribov, Lipatov, Altarelli, and Parisi

[5] when we repet this process using different sets of samples, the estimate should be centered on the true value, ie the estimator should be unbiased.

DIS data are insufficient to determine accurately many aspects of PDFs, such as the flavour decomposition of the quark and antiquark sea or the gluon distribution, especially at large x. Hence PDF determination must be based on global fits, in whch hadronic data are included along with DIS data.

CHAPTER 3

Relevant Statistics

In this section we provide only some of the basics and pre-requisites for statistical ideas that are relevant to our studies and to PDFs in general. Since these ideas are so central to studying PDFs, we only provide the minimum required knowledge here, and we shall return to more advanced statistics later on and throughout the paper.

3.1 General Methods

3.1.1 Probability Distributions

In studying particle physics, one must be very aware of two very important probability distributions: the Posson distribution and the Gaussian distribution. The Gaussian distribution is Ubiquotous in all fields and it is the most famous distribution, describing a wide range of phenomena. The Poisson distibution is a probability mass function, describing (discrete) count data. It has one parameter, λ, which is the expected value. It is given by

$$Pois(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{3.1}$$

It describes the probability of of getting x counts, where the mean count is λ.

The 1-D Gaussian is

$$\mathcal{N}\left(x \mid \mu, \sigma^2\right) = \frac{1}{\left(2\pi\sigma^2\right)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \tag{3.2}$$

The multivariate Gaussian for a D-dimenssional vector $\mathbf{x}$ is

$$\mathcal{N}(\mathbf{x} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \tag{3.3}$$

This can be viewed as each $x \in \mathbf{x}$ being normally-distributed with a mean $\mu \in \mu$ and standard deviation $\sigma \in \boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma}$ is the covariance matrix.

This can also be rewritten in terms of a χ^2, where $\chi^2(\vec{y},\vec{t})$ is defined in the standard way

$$\chi^2(\vec{y},\vec{t}) = (\vec{y}-\vec{t})^t\Sigma^{-1}(\vec{y}-\vec{t}) \tag{3.4}$$

So the conditional probability for new data to be confined in a differential volumt d^ny around $\vec{y}$ for a given configuration of parameters $\vec{\alpha}$ is

$$\mathcal{P}(\vec{y} \mid \vec{\alpha})d^ny = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}e^{-\frac{1}{2}\chi^2(\vec{y},\vec{t})}d^ny \tag{3.5}$$

Since we are interested in the modelling of data, we are interested in finding the set of fitted parameters that are most likely to be correct. In order to quantify the accuracy of the parameters, we identify the probability of the data given the parameters as the likelihood of the parameters, given the data.

For example, if I have a measurement that produces some data $\bar{x}$, then the *model* for the hypothesis is $P(\bar{x}|\theta)$, which depends on parameter θ. If I view this as just a function of the parameter, we supress $\bar{x}$ and call it the likelihood.

$$P(\bar{x}|\theta) = L(\theta) \tag{3.6}$$

The x is still there ("plugged in") but to quantify the parameters it is more convenient that we only look at the dependence on the parameter, constructing the likelihood $L(\theta)$.

3.1.2 Method of Least Squares

One of the simplest and most widely-used method for estimating the values of parameters is using the Method of least squares. Suppose you have independent variables x_i and dependent variables y_i that are found by ibservation. You construct a model $f(x;\theta)$ that aims to model y_i and is parameterized by θ. The goal is to find the parameter values θ that best fit the collected data. The fit of a model to data is found by its residuals, which are defined as the difference between the actual observed values and the model predictions for any given data point.

$$r_i = y_i - f(x_i, \beta) \tag{3.7}$$

You then calculate the sum of the least squares S

$$S = \sum_{i=1}^{n} r_i^2 \tag{3.8}$$

The values of the best-fit parameters are then found by minimizing S w.r.t. to the intended parameter. For ex, if you want to find $\hat{\theta}_1$ then solve for $\hat{\theta}_1$ using

$$\frac{\partial S}{\partial \theta_1}|_{\hat{\theta}_1 = \theta_1} = 0 \tag{3.9}$$

3.1.3 Maximum Likelihood Method

The maximum Likelihood method is a method for parameter estimation. It consists of finding the value(s) of θ which maximize the likelihood. This value is calked the estimator $\hat{\theta}$ (for the true value θ). It (they) is (are) determined by solving

$$L(\hat{\boldsymbol{\theta}}; \mathbf{X}) = \text{argmax}_{\boldsymbol{\theta}} L(\boldsymbol{\theta}; \mathbf{X}) \tag{3.10}$$

That means we just take the derivative w.r.t $L(\theta)$ set it to 0 and solve for $\theta = \hat{\theta}$

$$\left.\frac{\partial L(\boldsymbol{\theta}; \boldsymbol{x})}{\partial \theta_i}\right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = 0; \quad i = 1, \ldots, R \tag{3.11}$$

The nice properties of the MLE do not necessarily carry over small statistics cases. For example with $(x_1, \ldots, x_N)$ sampled from a normal distribution $N(\mu, \sigma^2)$ the MLEs for μ and σ^2 are

$$\begin{aligned} \hat{\mu} = \bar{x} &\equiv \frac{1}{N} \sum_{n=1}^{N} x_n \\ \hat{\sigma}^2 &= \frac{1}{N} \sum_{n=1}^{N} (x_n - \bar{x})^2 \end{aligned} \tag{3.12}$$

Notice that $\hat{\mu}$ is an unbiased estimator for μ (since $b(\hat{\mu}) =< \hat{\mu} > -\hat{\mu} = \bar{x} - \bar{x} = 0$), whereas $\hat{\sigma^2}$ has bias $b(\sigma^2) \equiv \langle\hat{\sigma}^2\rangle - \sigma^2 = -\frac{1}{N}\sigma^2$. Here it is easy to correct for, to obtain an unbiased estimator $\frac{N}{N-1}\hat{\sigma}^2$.

3.1.4 Method of Maximum Likelihood (Cowan)

The maximum likelihood estimator is the value of θ that maximizes the likelihood $\hat{\theta}_{MLE} = argmax_\theta L(\theta)$.

Finding the maximum is equivalent to finding the log of the likelihood $lnL(\theta)$, because log is a monotonic function.

Say we have data that is described by an exponential distribution $f(t;\tau) = \frac{a}{\tau}e^{-t/\tau}$. Suppose that we have n particle decays and for each one we observe the decay time τ, and suppose they're all iid, then each of these particles follows the same distribution, then the joint pdf of this entire dataset is... and if we view it as a function not of the data but of the parameter τ then

$$f(t_1,...,t_n|\tau)\prod_{o=1}^{n} f(t_i|\tau) = L(\tau) \tag{3.13}$$

$$lnL(\tau) = \sum_{i=1}^{n}(ln\frac{1}{\tau} - \frac{t}{\tau})$$

the derivative to zero

$$\frac{\partial L}{\partial\tau} = \sum_{i=1}^{n}(-\frac{1}{\tau} + \frac{t}{\tau}) \tag{3.14}$$

$$\hat{\tau}(t_1...t_n) = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{3.15}$$

From here we can examine this estimator, is it biased, and what is its variance?

If the t_is are all independent, then variance of the sum is the sum of the variances

$$V[\hat{\tau}] = V[\frac{1}{n}\sum_i t_i] = \frac{1}{n^2}\sum_i V[t_i] = \frac{\tau^2}{n} \tag{3.16}$$

$$\sigma = \frac{\tau}{\sqrt{n}}$$

Information inequality: $V[\hat{\theta}] \geq$ something.

Say we have the data $\vec{\theta}$, from which we constructed the estimator, from which we get the covariance matrix of each pair of estimators.

$$\text{covariance matrix} = V_{ij} = \tag{3.17}$$

The variance for an estimator $\hat{\theta}$ is

$$V[\hat{\theta}] = -\frac{1}{(\frac{\partial^2\ lnL}{partial\theta^2})|_{\theta=\hat{\theta}}} = \hat{\sigma}^2_{\hat{\theta}} \tag{3.18}$$

Expand $lnL(\theta)$ about $\hat{\theta}$

$$\log L(\theta) = \log L(\hat{\theta}) + \left[\frac{\partial \log L}{\partial \theta}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta}) + \frac{1}{2!} \underbrace{\left[\frac{\partial^2 \log L}{\partial \theta^2}\right]_{\theta=\hat{\theta}}}_{\sigma\hat{\theta}} (\theta - \hat{\theta})^2 + ... \tag{3.19}$$

$$\log L(\theta) = \log L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\widehat{\sigma^2}_{\hat{\theta}}} \tag{3.20}$$

If we plot lnL vs θ we get a bell distribution, the peak of the curve is $\hat{\theta}$. If we consider moving θ away from $\hat{\theta}$ by one standard deviation (one $\hat{\sigma}_{\hat{\theta}}$, then the log likelihood will decrease by 1/2 from its maximum value, and that distance will then give you the standard deviation of the estimator graphically.

In finite data samples, the log likelihood is usually not perfectly parabolic, so it might lean to one side more than the other, ie it has assymetric "widths" from its max. You can define the two widths as $\Delta\theta_+$ and $\Delta\theta_-$ and define the confidence interval

$$\text{confidence interval} = [\hat{\theta} - \Delta\hat{\theta}_-, \hat{\theta} + \Delta\hat{\theta}_+] \tag{3.21}$$

3.1.5 Chi-squared Test

The chi-squared is figure of merit (that measures the agreement between data and model predictions) and the chi-squared test is a goodness-of-fit test, which evaluates how well your probability model fits the observed data. Suppose you have observed data $\mathcal{O} = \{O_1, O_2, ..., O_N\}$ where i is a given observed value (for binned data set, for example, for N_i being the number of observations in the ith bin and probability model predictions $\mathcal{E} = \{E_1, E_2, ..., E_N\}$ (for a binned distribution $\mathcal{E}_\rangle = n_i$ is the number of events in the ith bin predicted from a known distribution. Note that for the binned example case, N_i must be an integer, whilst n_i may not be) and the uncertainties $\sigma = \{\sigma_1, \sigma_2, ..., \sigma_N\}$. The chi-squared parameter is

$$\chi^2 = \sum_i^N \frac{(o_i - E_i)^2}{\sigma_i^2} \tag{3.22}$$

[1]

Where $(o_i - E_i)^2$ are called the residuals. Clearly, you want χ^2 to be as low as possible for the best fit. However, it is possible for the χ^2 to be too low (overfit). The way in which we can determine how good our fit is, is using the reduced chi-square which is

$$\text{Reduced } \chi^2 = \frac{\chi^2}{\nu} \tag{3.23}$$

Where ν is the number of degrees of freedom, which is given by

$$\nu = N_{\text{observations}} - N_{\text{parameters in fit}} \tag{3.24}$$

For example, if you're using a straight line as your model to fit your data, which is given by $y = mx + b$, you have m and b as your model parameters, hence $N_{\text{parameters in fit}} = 2$. If you are using a sine wave, which is given by $y = Asin(\theta x + \phi)$, then A, θ and ϕ are your model parameters, hence $N_{\text{parameters in fit}} = 3$.

If the fit (model) is perfect, then Reduced $\chi^2 = 1$, which tells you that the scatter around your points, is about what you'd expect from the errors. The value of the Reduced $\chi^2 > 1$, indicates null hypothesis is rather unlikely, meaning the fit (model) could be improved. If then Reduced $\chi^2 < 1$ the fit (model) is too close to the observed data, i.e. you have too complicated of a model, or that your errors are too big. A good rule of thumb is that a "typical" value of χ^2 for a moderately good fit is $\chi^2 \approx \nu$, since the χ^2 statistic has a mean ν and standard deviation $\surd$.

[1] For a multivariate distribution, the chi-squared is given by $\chi^2(\vec{y}, \vec{t}) = (\vec{y} - \vec{t})^t \Sigma^{-1} (\vec{y} - \vec{t})$ where $\vec{y}$ is the observations, $\vec{t}$ are the model predictions, and Σ is the covariance matrix of the observations.

As we shall see later, the χ^2 is used as a figure of merit and is minimized to attain the parameter values and their error estimates in xFitter and other PDF libraries. It is important to underline the importance of the statistical measure of any goodness-of-fit estimator. If our measure suggests that the model is an unlikely match to the data, the parameters and their errors are basically worthless.

3.1.6 Confidence Intervals

Confidence intervals are sets covering the possibilities of multidimensional parameter spaces. The basic idea of confidence intervals is: for any possible θ we construct a set of observations x for which that value of θ will be included in the $1-\alpha$ region. We call this set $S_\alpha(\theta)$. There are several ways we can construct these intervals, in all we look for the smallest set for which

$$\int_{S_\alpha(\boldsymbol{\theta})} f(\boldsymbol{x};\boldsymbol{\theta})\mu(dS) \geq 1-\alpha \tag{3.25}$$

Where $\boldsymbol{x} \in S_\alpha(\boldsymbol{\theta})$. Given an observation x, the confidence interval $C_\alpha(x)$ for θ at the $1-\alpha$ confidence level is the set of all values of θ for which $x \in S_\alpha(\theta)$. This set has a probability $1-\alpha$ or more of including the true value of θ. We are trying to say the statement:

$$P(\theta_{true} \in region) \geq 1-\alpha \tag{3.26}$$

Where α is something small (like 5 %) and $1-\alpha$ is called the confidence limit (or credibility level).

Frequentest Approach If I have a parameter θ, I will construct a test. Ex. A Poisson counting experiment. Assume the data that we get is a number n

$$n\ Poisson(s+b) \tag{3.27}$$

Where s is the expected number of signal events and b is the expected number of background events.

Suppose that $n_i \sim Poisson(\mu s_i + b_i)$ where b_i and s_i are known constants from MC simulation. We want to get a t value for a given value of μ. One way of doing this.

is to generate a bunch of MC and calculate the integral. Another way that avoids doing this is Wilk's theorem. It says

$$f(t_\mu|\mu) \sim \underbrace{\chi_1^2}_{chi-square for\ n=1\ dof} \tag{3.28}$$

Hence the p-value is

$$p_\mu = 1 - F_{\chi_1^2}(t_\mu) \tag{3.29}$$

where $F_{\chi_1^2}(t_\mu)$ is the comulative dist. for the χ^2 with one dof. We set this equation to α

$$p_\mu = 1 - F_{\chi_1^2}(t_\mu) = \alpha \tag{3.30}$$

Hence

$$t_\mu = F_{\chi_1^2}^{-1}(1-\alpha) \tag{3.31}$$

Also,

$$t_\mu = -2\ln\frac{L(\mu)}{L(\hat{\mu})}, \text{ By def.} \tag{3.32}$$

Hence if we want to know the value of μ that defines the boundary of that region just equate 3.31 to 3.32 and solve for $L(\mu)$

$$\ln L(\mu) = \ln(\hat{\mu}) - \frac{1}{2}F_{\chi_1^2}^{-1}(1-\alpha) \tag{3.33}$$

Where $ln(\hat{\mu})$ is of course just the max of the log-likelihood function. So to find the bounds of the confidence interval, you first find the value of μ that maximizes the log-likelihood function, and then you move away from that until the log-likelihood decreases from its maximum by a certain amount $frac12F_{\chi_1^2}^{-1}(1-\alpha)$, where α is decided beforehand (for ex. α could by 5 % or something). For example, if I take

$$1-\alpha = \underbrace{68\%}_{a=1} \tag{3.34}$$

And let's define a number a to be the number of standard deviations, such that when I integrate a gaussian between those limits I get a certain $1-\alpha$

$$1-\alpha = \int_{-a}^{a} \phi(x)dx \tag{3.35}$$

Where $\phi(x)4$ is the standard normal distribution of x. So that defines the number of standard deviations of a gaussian between which I contain an area $1-\alpha$. What you can show is that if x is Gaussian distributed, then $x^2 \sim \chi_1^2$. And hence

$$1-\alpha = \int_{0}^{a} f_{\chi_1^2}(y)dy = F_{\chi_1^2}(a) \tag{3.36}$$

$$a = F_{\chi_1^2}(1-\alpha) \tag{3.37}$$

Hence

$$\ln L(\mu) = \ln L_{max} - \frac{1}{2}, \text{ for } 1-\alpha = 68\% \tag{3.38}$$

3.1.7 Multivariate and χ^2

Say that the likelihood for θ is multivariate normal, the likelihood function of a single observation is of the form

$$\begin{aligned} L(\boldsymbol{\theta}; x) &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}[x - g(\boldsymbol{\theta})]^T\boldsymbol{\Sigma}^{-1}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})]\right\} \\ &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}\chi^2\right\} \end{aligned} \tag{3.39}$$

Taking the log and dropping the const in the beginning (indep. of θ, we have

$$\begin{aligned} \log L(\boldsymbol{\theta}; \boldsymbol{x}) &= -\frac{1}{2}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})]^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})] \\ &= -\frac{1}{2}\chi^2 \end{aligned} \tag{3.40}$$

Thus, $-2\log L$ is precisely the χ^2 expression in the least squares method.
Proof: suppose we want to fit N data points (x_i, y_i) $i = 1, ..., N$ to a model that has M parameters θ_j, $j = 1, ..., M$ given by $y(x) = y(x; \vec{\theta})$. In order to find the best-fit values

for θ_j, we minimize the least-squares fit

$$\sum_{i=1}^{N}[y_i - y(x_i;\vec{\theta})]^2 \tag{3.41}$$

Now suppose we that each data point y_i has a measurement error that is independently random and distributed as a normal distribution around the model $y(x)$. Suppose further that the standard deviations σ of these normal distributions are the same for all points (and that there are no correlations). The the probability of the data set is the product of the probabilities of each point

$$L \propto \prod_{i=1}^{N}\left\{\exp\left[-\frac{1}{2}\left(\frac{y_i - y\left(x_i;\vec{\theta}\right)}{\sigma}\right)^2\right]\Delta y\right\} \tag{3.42}$$

Maximizing 3.42 is equivalent to maximizing its logarithm, or equivalently minimizing the negative of its logarithm:

$$\left[\sum_{i=1}^{N}\frac{\left[y_i - y\left(x_i;\vec{\theta}\right)\right]^2}{2\sigma^2}\right] - N\log\Delta y \tag{3.43}$$

since N, σ and Δy are all constants, minimizing 3.43 is equivalent to minimizing 3.41.

If each data point (x_i, y_i) has its own known standard deviation σ_i, then the χ^2 is

$$\chi^2 \equiv \sum_{i=1}^{N}\left(\frac{y_i - y\left(x_i;\theta_1\ldots\theta_M\right)}{\sigma_i}\right)^2 \tag{3.44}$$

Equation 3.44 can be minimized with respect to the parameter θ_k to obtain equations that must hold at the chi-squared minimum

$$\begin{aligned} 0 &= \frac{\partial\chi^2(x_i;\theta_1,...,\theta_k,...)}{\partial\theta_k} \\ &= \sum_{i=1}^{N}\left(\frac{y_i - y\left(x_i\right)}{\sigma_i^2}\right)\left(\frac{\partial y\left(x_i;\theta_1\ldots\theta_k\ldots\right)}{\partial\theta_k}\right) \qquad k = 1,\ldots,M \end{aligned} \tag{3.45}$$

A popular method for estimating the errors in the maximum likelihood method is to look for parameters $\theta_\pm$ for which

$$-2\Delta\log L \equiv -2\left[\log L\left(\boldsymbol{\theta}_\pm;\boldsymbol{x}\right) - \log L(\hat{\boldsymbol{\theta}};x)\right] = 1 \tag{3.46}$$

That is,

$$-2\left[\log L\left(\boldsymbol{\theta}_\pm;\boldsymbol{x}\right) - \log L(\hat{\boldsymbol{\theta}};x)\right] = \Delta\chi^2 = 1 \tag{3.47}$$

This method yields 68 $\%$ condidence intervals on the individual parameters as long as f_X is normal (in other words, this choice implies that the likelihood is multivariate Gaussian.) The basis of this assumption is a significant motivation of this study. More on this subject in Chapter 6 If the sampling is not normal, then the probability content will be different and must be determined for the correct sampling distribution.

This PDF fits that employ the ideal choice $\Delta^2 = 1$ are usually limited to a smaller set of data, while the global fits prefer to take $\Delta\chi^2 > 1$ to account for small inconsistencies among the data sets and to compensate for the parameterization bias.

The shape of the χ^2 function around its minimum can be used to determine the confidence intervals of the parameters $\bar{\theta}$. By varying on parameter θ_i around the minimum and then minimizing the χ^2 function around all the other $\bar{\theta}$ parameters, one finds the limits at which this **profiled** χ^2 exceeds the difference of the tolerance value T to the minimumS. As discussed earlier, for an optimal fit, one chooses tolerance $T = 1$, as in 3.47. This profiling technique is used to construct the confidence intervals for the parameters. For a two-dimensional parameter space, see the figure 3.1.

3.2 Bayesian Reweighting

in a global fit, the goal is constructing a probability model, where a fitting technique such as χ^2 minization is used to determine (the parameters of) the probability density. This process is iterative: once new (experimental) data are available, a new fit is performed. It is desirable to not have to do a new global fit every time new data becomes available. Although a different reweighting method has been suggested recently based on the Hessian approach in "PDF reweighting in the Hessian matrix approach" , such updating can be done as a Bayesian inference problem: Bayesian reweighting.

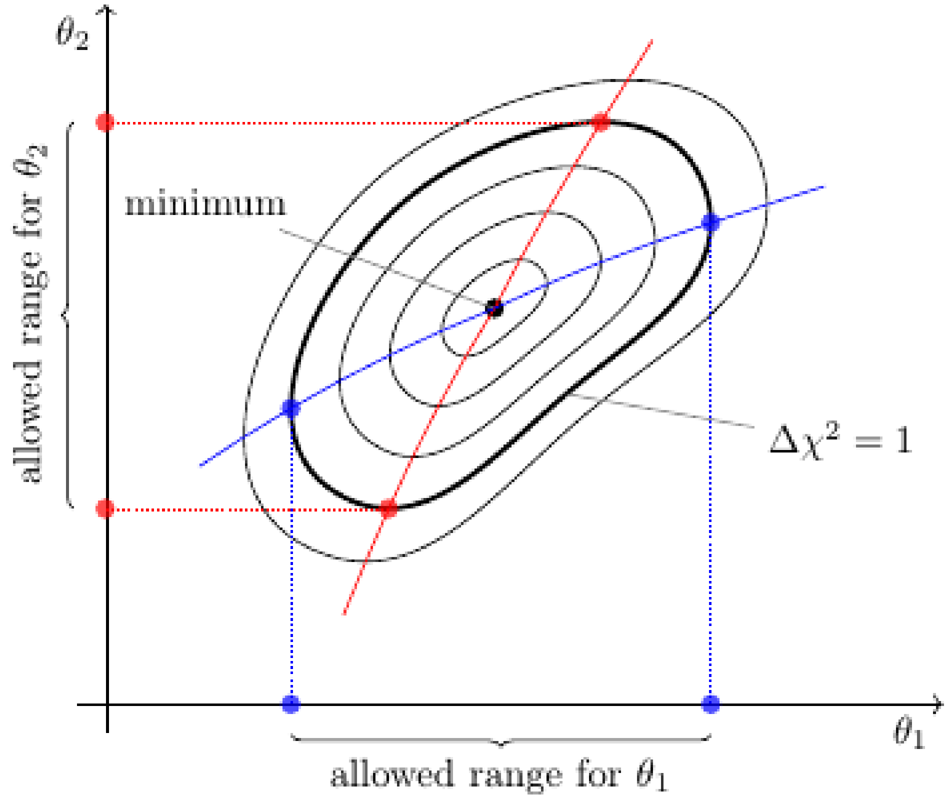


Fig. 3.1: Caption

Starting from Baye's theorem

$$\mathcal{P}(\vec{\alpha} \mid D) = \frac{\mathcal{P}(D \mid \vec{\alpha})}{\mathcal{P}(D)} \mathcal{P}(\vec{\alpha}) \tag{3.48}$$

Where $\mathcal{P}(\vec{\alpha} \mid D)$ is the posterior, $\mathcal{P}(D \mid \vec{\alpha})$ is the likelihood, which represents the conditional probability for a dataset D, given the parameters $\bar{\alpha}$ of the model. For an observable $\mathcal{O}$, the expectation value of this observable can be written as

$$\begin{aligned} \mathrm{F}[\mathcal{O}] &= \int d^n\alpha \mathcal{P}(\vec{\alpha} \mid D)\mathcal{O}(\vec{\alpha}) \\ &= \int d^n\alpha \frac{\mathcal{P}(D \mid \vec{\alpha})}{\mathcal{P}(D)} \mathcal{P}(\vec{\alpha})\mathcal{O}(\vec{\alpha}) \\ &= \frac{1}{N}\sum_k w_k \mathcal{O}\left(\vec{\alpha}_k\right). \end{aligned} \tag{3.49}$$

Where in the last line, we used a Monte Carlo approximation [2] of the integral. The parameters $\{w_k\}$ are weights [3] that are propotional to $\mathcal{P}(D \mid \vec{\alpha})$, and with normalization fixed by $\sum_k w_k = N$.

The method of Bayesian reweighting breaks down into two different approachesL the Geile and Keller (GK) and the chi-square approach.

3.2.1 Geile and Keller (GK) Method

In a global fit, the underling probability distribution of of PDFs $\mathcal{P}_{old}(f)$ is represented by a large ensemble of PDFs $f_k, k = 1...N_{rep}$, where N_{rep} is the number of PDF replicas (as discussed earlier)

$$\mathcal{P}_{old} = \mathcal{P}_{old}(f_k); \quad k = 1...N_{rep} \tag{3.50}$$

[2] Monte carlo method tells us that the integral

$$\mathcal{F}(\boldsymbol{\theta}) = \int f(\mathbf{x})p(\mathbf{x} \mid \boldsymbol{\theta})d\mathbf{x} = \mathbb{E}_{p(x|\theta)}[f(\mathbf{x})]$$

can be approximated as a sum instead of an integral, by drawing samples from the distribution p, evaluate the function f at these samples, then compute the average. You then attain the estimator $\hat{\mathcal{F}}$ as

$$\hat{\mathcal{F}}(\boldsymbol{\theta}) = \frac{1}{N}\sum^{N} f\left(\hat{\mathbf{x}}^n\right); \quad \hat{\mathbf{x}}^n \sim p(\mathbf{x} \mid \boldsymbol{\theta}), \text{ for } n = 1, \ldots, N$$

[3] such as the weights in importance sampling

One can compute the expectation value of an observable $\mathcal{O}$

$$\mathrm{E}[\mathcal{O}]=\frac{1}{N}\sum_{k} w_k \mathcal{O}\left(\vec{\alpha}_k\right) \tag{3.51}$$

and the variance

$$\operatorname{Var}[\mathcal{O}]=\sqrt{\frac{1}{N_{\text{rep}}}\sum_{k=1}^{N_{\text{rep}}}\left(\mathcal{O}\left[f_k\right]-\langle\mathcal{O}\rangle\right)^2}. \tag{3.52}$$

According to Baye's theorem in 3.49 the initial probability distribution $\mathcal{P}_{\updownarrow\lceil}(\{)$ can be updated to also include additional (new) data set $\vec{y}$

$$\mathcal{P}_{\text{new}}\left(f\right)\propto\mathcal{P}(\vec{y}\mid f)\mathcal{P}_{\text{old}}\left(f\right) \tag{3.53}$$

Where $\mathcal{P}(\vec{y}\mid f)$ is the conditional probability of new data given the set of PDFs. It follows that the average value of any observable depending on the PDFs becomes a weighted average

$$\begin{aligned}\langle\mathcal{O}\rangle_{\text{new}} &=\frac{1}{N_{\text{rep}}}\sum_{k=1}^{N_{\text{rep}}}\omega_k\mathcal{O}\left[f_k\right]\\ \operatorname{Var}[\langle\mathcal{O}\rangle_{\text{new}}] &=\sqrt{\frac{1}{N_{\text{rep}}}\sum_{k=1}^{N_{\text{rep}}}\omega_k\left(\mathcal{O}\left[f_k\right]-\langle\mathcal{O}\rangle_{\text{new}}\right)^2}\end{aligned} \tag{3.54}$$

Giele and Keller (GK) suggested that the likelihood function that one should use follows from taking $\mathcal{P}(\vec{y}\mid f)d^n y$ as the probability to fund new data to be confined in a differential element $d^n y$ around $\vec{y}$ resulting in

$$\omega_k^{\mathrm{GK}}=\frac{\exp\left[-\chi_k^2/2\right]}{\left(1/N_{\text{rep}}\right)\sum_{k=1}^{N_{\text{rep}}}\exp\left[-\chi_k^2/2\right]} \tag{3.55}$$

Where

$$\chi_k^2=\sum_{i,j=1}^{N_{\text{data}}^{\text{New}}}\left(y_i\left[f_k\right]-y_i\right)C_{ij}^{-1}\left(y_j\left[f_k\right]-y_j\right) \tag{3.56}$$

3.2.2 chi-squared

The other option for determining the likelihood function, advocated by the NNPDF collaboration derives from taking $\mathcal{P}(\vec{y} \mid f)d_\chi$ as the probability density for the corresponding chi-squared value $\chi = \sqrt{\chi^2}$ to be confined in a differential volume $d\chi$ around χ instead

$$\omega_k^{\text{chi-squared}} = \frac{(\chi_k^2)^{(N_{\text{data}}-1)/2} \exp\left[-\chi_k^2/2\right]}{(1/N_{\text{rep}}) \sum_{k=1}^{N_{\text{rep}}} (\chi_k^2)^{(N_{\text{data}}-1)/2} \exp\left[-\chi_k^2/2\right]} \tag{3.57}$$

It was pointed out by Prosper that the GK weights w^{GK} carry more information than the chi-squared weights $w^{\text{chi-squared}}$, as a given data set uniquely determines the value of χ^2, while a fixed χ^2 may correspond to different data set.

[4]

The ensemble of PDFs required by the Bayesian approach can be constructed as

$$f_k \equiv f_{S_0} + \sum_i^{N_{\text{eig}}} \left(\frac{f_{S_i^+} - f_{S_i^-}}{2}\right) R_{ik} \tag{3.60}$$

Where the coeffiecients Rik are random numbers draw from a Gaussian distribution centered at zero with variance one. An assymetric version of the PDF replicas was suggested as

$$f_k^{\text{asym}} \equiv f_{S_0} + \sum_i^{\text{Nerg}} \left(f_{S_i^\pm} - f_{S_0}\right) |R_{ik}| \tag{3.61}$$

this assymetric version was suggested to account for non-linearities. After computing the weights w_k for each replica, the reweighted PDFs can be written as

$$f_{\text{new}} = f_{S_0} + \sum_i^{N_{\text{eig}}} \left(\frac{f_{S_i^+} - f_{S_i^-}}{2}\right) \left(\frac{1}{N_{\text{rep}}} \sum_k^{N_{\text{rep}}} \omega_k R_{ik}\right), \tag{3.62}$$

[4] In summary, one has two options to write the weights following two different probability densities for new data, the option suggested by Giele and Keller

$$\mathcal{P}(\vec{y} \mid \vec{y}) = \frac{\mathcal{P}(\vec{y} \mid \vec{y})}{\mathcal{P}(\vec{y})}\mathcal{P}(\vec{y}) \quad \rightarrow \quad w_k^{GK} \propto \exp\left(\frac{1}{2}\chi^2\left(\vec{y}, \overrightarrow{t_k}\right)\right) \tag{3.58}$$

or the one suggested by the NNPDF collaboration:

$$\mathcal{P}(\vec{y} \mid \chi) = \frac{\mathcal{P}(\chi \mid \vec{y})}{\mathcal{P}(\chi)}\mathcal{P}(\vec{y}) \quad \rightarrow \quad w_k^{\text{chi-squared}} \propto \left(\chi^2\left(\vec{y}, \vec{t}_k\right)\right)^{\frac{1}{2}(n-1)} \exp\left(\frac{1}{2}\chi^2\left(\vec{y}, \vec{t}_k\right)\right) \tag{3.59}$$

And one can calculated the "penalty" induced in the original fit by

$$P = \Delta\chi^2 \sum_{i}^{N_{\text{eig}}} \left(\frac{1}{N_{\text{rep}}} \sum_{k}^{N_{\text{rep}}} \omega_k R_{ik} \right)^2 \tag{3.63}$$

3.3 The Hessian method

There are two established methods for error analysis: the Hessian method (described in detail elsewhere in this paper) or the Monte Carlo (MC) method. xFitter implements the Hessian method, which assumes a quadratic approximation of the function

$$\chi^2 = \chi_0^2 + \Delta\chi^2 \tag{3.64}$$

around its global minimum. Note that the χ^2 is a *global* goodness-of-fit quantity, so equation 3.64 can be written as

$$\Delta\chi^2_{\text{global}} \equiv \chi^2_{\text{global}} - \chi_0^2 \tag{3.65}$$

Here, χ_0^2 is the value of the function at the global minimum with the best-fit values $\{\theta_0\}$. For example,

$$\chi_0^2 = \sum_i \frac{\left(\mu_i - m(\{\hat{\theta_0}\})_i\right)^2}{\Delta_i^2} + \sum_\alpha b_\alpha^2 \tag{3.66}$$

and $\Delta\chi^2$ is the displacement from the minimum. The Hessian matrix H is constructed from the second derivatives of χ^2 at the minimum. The matrix H_{ij} is defined as

$$H_{ij} = \frac{1}{2} \left(\frac{\partial^2 \chi^2}{\partial y_i \partial y_j} \right) \Bigg|_{\text{min}} \tag{3.67}$$

With y_i begin the displacement of the parameter θ_i from its value θ_0 at the minimum

$$y_i = \theta_i - \theta_0 \tag{3.68}$$

For the analyzed function χ^2 one writes

$$\chi^2 = \chi_0^2 + \underbrace{\sum_{i,j} H_{ij} y_i y_j}_{=\Delta\chi^2} \tag{3.69}$$

Or, equivalently,

$$\Delta\chi^2_{\text{global}} \equiv \chi^2_{\text{global}} - \chi_0^2 = \sum_{i,j=1}^{n} H_{ij} \left(\theta_i - \theta_i^0\right) \left(\theta_j - \theta_j^0\right) \tag{3.70}$$

The nice thing about the Hessian method is that the uncertainty on a quanitity $F\left(\{a_i\}\right)$ is then obtained from linear error propagation

$$\Delta F = T \sqrt{\sum_{i,j=1}^{n} \frac{\partial F}{\partial a_i} C_{ij} \frac{\partial F}{\partial a_j}} \tag{3.71}$$

Where C is the covariance matrix, and $T = \left(\Delta\chi^2_{\text{global}}\right)^{1/2}$ is the tolerance for the reqiured vconfidence interval.

The Hessian (or covariance) matrix is symmetric and has a complete set of orthonormal eigenvectors ν_{ij}. It more convenient to diagonaliza this matrix and work with its orthonormal eigenvectors v_k

$$\sum_{j=1}^{n} C_{ij} v_{jk} = \epsilon k v_{ik} \tag{3.72}$$

where ϵ_k is the kth eigenvalue and v_{ik} is the ith component of the kth orthonormal eigenvector $(k = 1, ..., n)$. The parameter displacements $Y - i$ can be expanded in this basis

$$y_i = \sum_j v_{ij} \sqrt{\frac{1}{\epsilon_j}} z_j \tag{3.73}$$

defining the rescaled eigenvectors $e_{ik} = \sqrt{\epsilon_k} v_{ik}$ we get

$$\theta_i - \theta_i^0 = \sum_{k=1}^{n} e_{ik} z_k \tag{3.74}$$

leading to the simplified relation

$$\Delta\chi^2 = \chi^2 - \chi_0^2 = \sum_i z_i^2 \tag{3.75}$$

That is, $\sum_{k=1}^{n} z_k^2 \leq T^2$ is the interior of a hypersphere of radius T. The varied parameters θ_i from which the resulting error sets are defined can be written as

$$\theta_i = \theta_0 \pm \Delta\theta_i = \theta_0 \pm \Delta\chi^2 \sum_j \frac{v_{ij}^2}{\epsilon_j} \tag{3.76}$$

The uncertainties for a given observable (or a quantity) $\mathcal{O}$, which can be a PDF or a derived quantity such as a cross section, can be calculated via

$$\Delta\mathcal{O} = \frac{1}{2}\sqrt{\sum_{k=1}^{n}\left[\mathcal{O}\left(S_k^+\right) - \mathcal{O}\left(S_k^-\right)\right]^2} \tag{3.77}$$

Where $\mathcal{O}(S_0)$ is the observable calculated with the central PDF set and the $S_i^{\pm}$ correspond to the error sets in the positive and negative directions calculated from the diagonalized parameter z_i.

Assymeteric errors can also be calculated with

$$\begin{aligned}(\Delta\mathcal{O})_+ &= \sqrt{\textstyle\sum_{k=1}^{n}\left\{\max\left[\mathcal{O}\left(S_k^+\right) - \mathcal{O}\left(S_0\right), \mathcal{O}\left(S_k^-\right) - \mathcal{O}\left(S_0\right), 0\right]\right\}^2}, \\ (\Delta\mathcal{O})_- &= \sqrt{\textstyle\sum_{k=1}^{n}\left\{\max\left[\mathcal{O}\left(S_0\right) - \mathcal{O}\left(S_k^+\right), \mathcal{O}\left(S_0\right) - \mathcal{O}\left(S_k^-\right), 0\right]\right\}^2}.\end{aligned} \tag{3.78}$$

In the ideal case one would choose the tolerance criterion so that $\Delta\chi^2 = 1$ (that represents the best fitting). But since we consider several data sets which are not necessarily in mutual agreeement with each other and have different experimental errors, such a choice would under-estimate the true underlying uncertainty (surely the actual errors are higher than if we were to consider all the experimental data on an equal footing). For the proton baseline with 13 free fit parameters it becomes $\Delta\chi^2 = 20$ at the 90 % confidence level. The choice of these values is determined from "EPPS16: Nuclear parton distributions with LHC data" or "Determination of nuclear parton distribution functions and their uncertainties at next-to-leading order". **This choice for the** $\Delta\chi^2$

tolerance is one of the main issues we are studying in this paper. Much more has been written on "Dynamic Tolerance", such as in the "PDFs for the LHC" paper.

The Hessian approach is based on a quadratic approximation to χ^2 global in the neighborhood of the minimum that defines the best fit. It yields a set of PDFs associated with the eigenvectors of the Hessian, which characterize the PDF parameter space in the neighborhood of the global minimum in a process-independent way.

The theory contains free parameters $\{a_i\} = \{a_1, \ldots, a_d\}$. w that characterize the nonperturbative input to the analysis. Fitting theory to experiment determines these parameters and thereby the PDFs. The uncertainty of the result due to experimental and theoretical errors is assessed in our analysis by an assumption on the permissible range of $\Delta\chi^2$ The analysis is based on an effective global chi-squared function that measures the quality of the fit between theory and experiment:

$$\chi^2_{\text{global}} = \sum_n w_n \chi^2_n \tag{3.79}$$

where n labels the 15 different data sets. The weight factors wn in (3), with default value 1, are a generalization of the selection process that must begin any global analysis, where one decides which data sets to include (w = 1) or exclude (w = 0).

The generic form for the individual contributions

$$\chi^2_n = \sum_I \left(\frac{D_{nI} - T_{nI}}{\sigma_{nI}}\right)^2 \tag{3.80}$$

where T_{nI}, D_{nI}, and σ_{nI} are the theory value, data value, and uncertainty for data point I of data set (or “experiment”) n. In practice, Eq. 3.80 is generalized to include correlated errors such as overall normalization factors; or even the full experimental error correlation matrix if it is available.

Having specified the effective 2 function, we find the parameter set that minimizes it to obtain a “best estimate” of the true PDFs. This PDF set is denoted by S0. The parameters that specify S_0 are noted in a table.

The most efficient approach to studying uncertainties in a global analysis of data is through a quadratic expansion of the 2 function about its global minimum.b This is

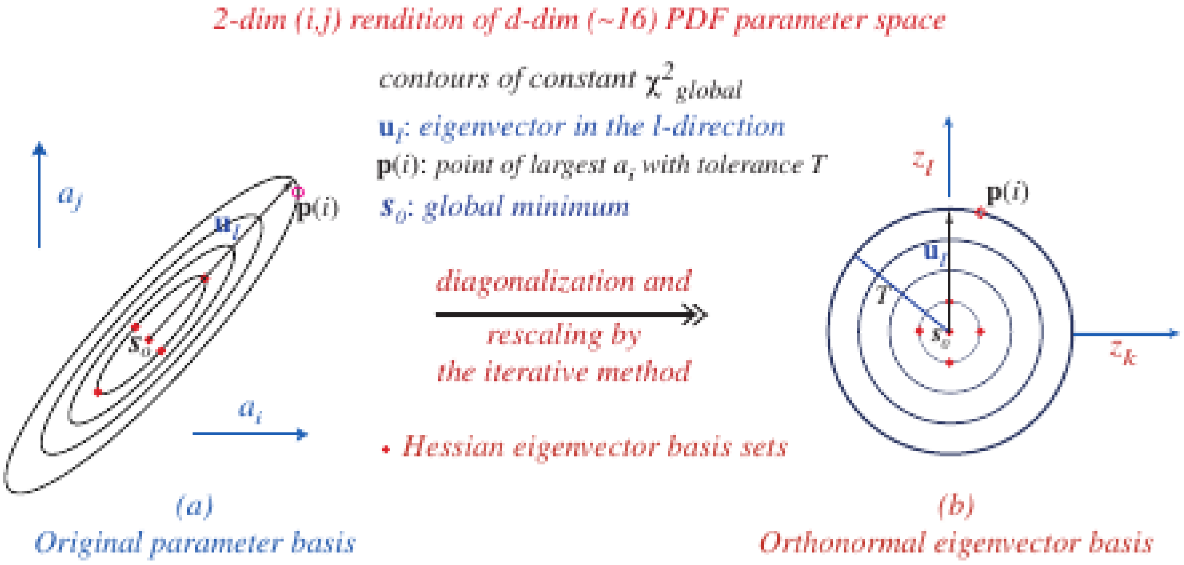

2-dim (i,j) rendition of d-dim (~16) PDF parameter space
contours of constant χ² global
u_l: eigenvector in the l-direction
p(i): point of largest a_i with tolerance T
s_0: global minimum
diagonalization and
rescaling by
the iterative method
Hessian eigenvector basis sets
a_j
a_i
u_l
p(i)
s_0
z_l
z_k
T
(a)
Original parameter basis
(b)
Orthonormal eigenvector basis

Fig. 3.2: Visualizing the Hessian method: the Hessian matrix is diagonalized and its eigenvalues are rescaled

the well known Error Matrix or Hessian method. The Hessian matrix is the matrix of second derivatives of χ^2 at the minimum.

The standard error matrix approach begins with a Taylor series expansion of 2 global(S) around its minimum S0, keeping only the leading terms. This produces a quadratic form in the displacements from the minimum:

$$\Delta\chi^2_{global} = \chi^2_{global} - \chi^2_0 = \sum_{i=1}^{d}\sum_{j=1}^{d} H_{ij}\left(a_i - a_i^0\right)\left(a_j - a_j^0\right) \tag{3.81}$$

The general idea is illustrated in figure 3.2, which is taken from "Uncertainties of predictions from parton distribution functions II: the Hessian method" The uncertainty on parameter ai in the global analysis follows from the master equation (24):

$$\Delta a_i = T\left(\sum_k M_{ik}^2\right)^{1/2} \tag{3.82}$$

The uncertainty range of the PDFs themselves can also be explored using the eigenvector method. For example, letting the gluon distribution g(x, Q) at some specific values of x and Q be the variable X that is extremized by the method of Sec. 3.4 leads to the extreme gluon distributions shown in the left-hand side of Fig. 3. The envelope of such curves, obtained by extremizing at a variety of x values at fixed Q, is shown by the shaded region, which is defined by T =10, i.e., by allowing 2 global up to 100 above its minimum value.

3.3.1 The Hessian Method written another way

The usual definition of the optimal correspondance between data and a set of PDFs $f = f(x, Q^2)$ that depends on the fit parameters $\{a\}$ is the minimum of a χ^2 function. In the most simple form we can write it as

$$\chi^2\{a\} = \sum_k \left[\frac{X_k^{\text{theory}}[f] - X_k^{\text{data}}}{\delta_k^{\text{data}}} \right]^2 \tag{3.83}$$

Where $X_k^{\text{theory}}[f]$ are the theory predictions depending on the PDFs, and δ_k^{data} is the corresponding uncertainty from data. In the Hessian approach, to quantify the PDF errors, the behavior of the χ^2 is expanded around the best fit S_0 by a second order polynomial in the space of fit parameters $\{a\}$

$$\chi^2\{a\} \approx \chi_0^2 + \sum_{ij} \delta a_i H_{ij} \delta a_j \tag{3.84}$$

Where χ_0^2 is the minimum value of χ^2 and $\delta a_j \equiv a_j - a_j^0 a$ are the excursions from the best-fit values. Note, however, that this choice, implies that the likelihood is multivariate Gaussian. This assumption of normal sampling is often forgotten.

This PDF fits that employ the ideal choice $\Delta^2 = 1$ are usually limited to a smaller set of data, while the global fits prefer to take $\Delta\chi^2 > 1$ to account for small inconsistencies among the data sets and to compensate for the parameterization bias.

Being symmetric, the Hessian matrix H_{ij} has N_{eig} orthonormal sets of eigenvalues ϵ_k and $v^{(k)}$ eigenvectors satisfying

$$\begin{gathered} H_{ij} v_j^{(k)} = \epsilon_k v_i^{(k)} \\ \sum_j v_j^{(k)} v_j^{(\ell)} = \sum_j v_k^{(j)} v_\ell^{(j)} = \delta_{k\ell} \end{gathered} \tag{3.85}$$

Defining a new set of variables as

$$z_k \equiv \sqrt{\epsilon_k} \sum_j v_j^{(k)} \delta a_j \tag{3.86}$$

one easily finds that

$$\chi^2\{a\} \approx \chi_0^2 + \sum z_i^2 \tag{3.87}$$

That is, this transformation diagonalizes the Hessian matrix. A criterion is neede to specify how much the term $\sum_i z_i^2$ can grow while the corresponding PDFs still remain "acceptable". It follows that the corresponding uncertainty for a PDF dependent quantity $O = O(f)$ can be computed as

$$(\Delta\mathcal{O})^2 = \Delta\chi^2 \sum_k \left(\frac{\partial\mathcal{O}}{\partial z_k}\right)^2 \tag{3.88}$$

An essential feature of the Hessian approach is the introduction of the PDF error sets S ± k , defined customarily (along with the best fit S0) in the z-space as

$$\begin{aligned} z\left(S_0\right) &= (0,0,\ldots,0), \\ z\left(S_1^{\pm}\right) &= \pm\sqrt{\Delta\chi^2}(1,0,\ldots,0) \\ z\left(S_2^{\pm}\right) &= \pm\sqrt{\Delta\chi^2}(0,1,\ldots,0) \\ &\vdots \\ z\left(S_{N_{\text{eig}}}^{\pm}\right) &= \pm\sqrt{\Delta\chi^2}(0,0,\ldots,1). \end{aligned} \tag{3.89}$$

One can thus calculate the derivatives as

$$\left(\frac{\partial\mathcal{O}}{\partial z_k}\right) \approx \frac{\mathcal{O}\left[S_k^+\right] - \mathcal{O}\left[S_k^-\right]}{2\sqrt{\Delta\chi^2}} \tag{3.90}$$

such that

$$(\Delta\mathcal{O})^2 = \frac{1}{4}\sum_k \left(\mathcal{O}\left[S_k^+\right] - \mathcal{O}\left[S_k^-\right]\right)^2 \tag{3.91}$$

3.4 The Bootstrap

[5]

3.5 Closure Test

How do we know that we have the right and faithful PDFs uncertainties? The problem is complicated since the true PDFs are not known. The idea to answer this is the so-called closure testing.

A closure test aims to evaluate only the fitting methodology, and make sure that the choice of different functional forms independently yields the same result. What is done is that a particular "truth" or "correct answer" is assumed apriori, allowing us to directly evaluate how accurate the relevant fits are.

One way this is performed is that a "true" PDF set (or fully defined parameterization and parameter values) is assumed (either through the functional form or through LHAPDF grids). Then, sets of pseudo-data are generated (for example randomly sampled from the values and covariance matrix from the "truth") based on this truth (For PDF studies, this pseudo-data should be cross sections), the fitting methodology is applied to this pseudo-data (i.e. for PDF studies, the fitting is done on this cross section pseudo-data instead of experimental cross section data), then results are finally compared to the underlying truth.

The independence of result on the particular choice of underlying truth can be explicitly tested by repeating the procedure with a different choice for the underlying "true" PDF.

Besides validating the particular fitting methodology, closure tests also allow for uncertainty estimation in a controlled manner. This is the primary aim of this study: how to validate or justify the uncertainty estimates on the PDF sets that are presented by groups such as the NNPDF collaboration. More on how this test is performed by the NNPDF collaboration is discussed on the NNPDF chapter.

[5] in quantum theory and quantum field theory, the bootstrap model, which was popular in the 1960s and 1970s, and which is facing a resurgence in its popularity in QFT, has a somewhat different meaning...

3.6 To Do

- Run xFitter on each of the datasets to get best fit values of the PDF parameters for each of these datasets. Question: Are these the parameters of the PDF functional form, for example, the HERAPDF PDF parameterizaion is

  $$Ax^B(1-x)^C\left(1+Dx+Ex^2\right)-A'x^{B'}(1-x)^{C'} \tag{3.92}$$

  ? And these are provided by xfitter after running it.

3.7 Notes on Papers

3.7.1 Parton Distributions with Theory Uncertainties: General Formalism and First Phenomenological Studies

issue of choosing a meaningful way to estimate the MHOU, which in particular incorporates these correlations. The standard way of estimating MHOUs in perturbative QCD calculations is to perform a variation of the renormalization and factorization scales with various choices for the range and combination of variations existing

3.7.2 A First Determination of Parton Distributions with Theoretical Uncertainties

At present, these uncertainties have two main origins. The first is the missing higher order uncertainty (MHOU) from the truncation of the QCD perturbative expansion. The second is related to knowledge of the structure of the colliding protons, as encoded in the parton distributions (PDFs)

Currently, PDF uncertainties only account for the propagated statistical and systematic errors on the measurements used in their determination.

3.7.3 Parton distributions for the LHC

This is where the NNPDF get the functional forms for the PDFs for the closure testing.

3.7.4 Parton distributions for the LHC Run II

This is where further tests, given the closure test data, are performed

3.7.5 xFitter 2.0.0: An Open Source QCD Fit Framework

3.7.6 Uncertainties of predictions from parton distribution functions II: the Hessian method

In the conventional approach, specific PDF sets are constructed to represent the "best estimate" under various input assumptions, including selective variations of some of the parameters. From these results, however, it is impossible to reliably assess the uncertainties of the PDFs or, more importantly, of the physics predictions based on them. The need to quantify the uncertainties for precision SM studies and New Physics searches in the next generation of collider experiments has stimulated much interest in developing new approaches to this problem.

The task is difficult because of the diverse sources of experimental and theoretical uncertainty in the global QCD analysis.

3.7.7 Closure Testing the NNPDF3.0 methodology

In a normal PDF fit to real experimental data, the underlying law which we are trying to estimate is unknown. This makes it difficult to evaluate how well a fitting methodology can reproduce the 'correct' answer and to ascertain whether there are sources of bias. In addition, all fits are to the same experimental data, so it is unclear whether an improvement in the quality of the fit to these data represents an actual improvement, or if it is due to over-learning (over-fitting) to the particular experimental dataset. Closure tests provide a way to evaluate the fitting methodology with a means to avoid these issues. Closure test works as follows:

first, we take the real data and replace

3.7.8 Bayesian Reweighting for Global Fits - Prosper

He generates datasets, and then divides them into 11 equally spaced regions in x and labeled as $\{d_0, d_1, \ldots, d_{10}\}$, then global fits are performed using χ^2 minimization to each data set, and then the parameters $\vec{\alpha}$ are obtained as $\vec{\alpha}_j^{\pm} = \vec{\alpha}_0 \pm \delta\vec{\alpha}_j$, then for reweighting, a monte carlo representation of fitted results is made by sampling the parameters as $\vec{\alpha}_k = \overrightarrow{\alpha_0} + \sum_j \delta\vec{\alpha}_j R_{kj}$ where R_{ij} are standard-normally distributed numbers. Then evaluating

$$f(x, \vec{\alpha}) = x^{\alpha_0}(1-x)^{\alpha_1}, \tag{3.93}$$

with parameters $\vec{\alpha_k}$ from each fit A_i gives the desired monte carlo sample $\{f_k \mid A_i\}$.

Then, the monte carlo sample $\{f_k \mid A_0\}$ is selected as the prior to be reweighted. Then, new data sets are $(B_i\}$ and they are used as new evidence, and the following are calculated: the expectation value $\mathrm{E}\left[f \mid A_0, B_i\right]$ and variance $\mathrm{Var}\left[f \mid A_0, B_i\right]$ using

$$\mathrm{E}[\mathcal{O}] = \frac{1}{N}\sum_k w_k \mathcal{O}\left(\vec{\alpha}_k\right) \tag{3.94}$$

where the weights are given by

$$\mathcal{P}(\vec{\alpha} \mid \vec{y}) = \frac{\mathcal{P}(\vec{y} \mid \vec{\alpha})}{\mathcal{P}(\vec{y})}\mathcal{P}(\vec{\alpha}) \quad \rightarrow \quad w_k \propto \exp\left(\frac{1}{2}\chi^2\left(\vec{y}, \overrightarrow{t_k}\right)\right) \tag{3.95}$$

or

$$\mathcal{P}(\vec{\alpha} \mid \chi) = \frac{\mathcal{P}(\chi \mid \vec{\alpha})}{\mathcal{P}(\chi)}\mathcal{P}(\vec{\alpha}) \quad \rightarrow \quad w_k \propto \left(\chi^2\left(\vec{y}, \vec{t_k}\right)\right)^{\frac{1}{2}(n-1)} \exp\left(\frac{1}{2}\chi^2\left(\vec{y}, \vec{t_k}\right)\right) \tag{3.96}$$

for each data set B_i. There is clear disagreement between using the two reweighting methods 3.96 and 3.95; the variances using 3.95 are greater than those in 3.96; and 3.96 is more compatible with global fits.

3.7.9 Updating and Optimizing Error PDFs in the Hessian Approach

The two most commonly-used methods for obtaining the PDF uncertainties are the Monte Carlo method and the Hessian method. In the Hessian method, a smaller number of sets than the MC method is needed. These "error sets" are used to obtain an estimate of the error. These error sets correspond to the plus and minus eigenvecor directions in the space of PDF parameters, which are used to approximate the χ^2 function near the global minimum. It is also possible to estimate the impact of new data directly using Hessian PDFs, as has been shown by Paukkunen and Zurita.

3.7.10 A new simple PDF parametrization: improved description of the HERA data

3.7.11 PDF reweighting in the Hessian matrix approach - 2014

In essence, the underlying probability distribution, represented in the NNPDF philosophy [6] by an ensemble (1000) of PDF replicas, is updated by assigning each replica a certain weight based on the new data. This method has become an increasingly popular way to estimate the effects of e.g. new LHC measurements [10–16]. The drawback is that it has been proven to work only in conjunction with the NNPDF fits while the majority of the existing PDF fits use a rather different way of quantifying the PDFs and their uncertainties. Along with the best fit found by 2 minimization, they provide a collection (50) of Hessian error sets [17] that quantify the neighborhood of the central fit within a certain confidence criterion $\Delta\chi^2$ 2 . An extension of the Bayesian reweighting technique to this particular case was suggested in [18], and has thereafter been used in some occasions [11, 19, 20]. However, a recent study [21] revealed clear deviations when comparing the results from reweighting to the ones obtained by a direct fit, indicating that the proposed generalization is not accurate.

NNPDF uses replicas of sets, with each given a certain weight (a la Bayesian reweighting). This has some drawbacks: it only works with the NNPDF sets, while different PDF fits use different ways of quantifying the PDFs and their uncertainties. Along with the best fit found by χ^2 minimization, they provide a collection(50) of Hessian error sets that quantify the neighborhood of the central fit within a certain confidence criterion $\Delta\chi^2$ 2

3.7.12 Open-source QCD analysis of nuclear parton distribution functions at NLO and NNLO

CHAPTER 4

Problems in PDFs: uncertainties, extrapolation, MHOTs,. . .

4.1 Problems of Uncertainty

Other groups have devised different parameterizations for the PDF than that in 2.11. For example, CTEQ 18 have the parameterization for the gluon-gluon PDF is:

$$g\left(x, Q = Q_0\right) = x^{a_1 - 1}(1-x)^{a2}\left[a_3(1-y)^3 + a_4 3y(1-y)^2 + a_5 3y^2(1-y) + y^3\right] \tag{4.1}$$

However, what is curious is that even though more parameters (and more data) were added in CTEQ18 vs CTEQ14, the error bands were larger on these new sets. This leads to the unusual situation where adding more parameters and more data leads to larger uncertainties, and it means that the uncertainties were underestimated by the bias, which was implicit in the parameterization.

[1] uncertainties on the fit parameters determined by least-squares and standard error propagation turned out to be smaller by about one order of magnitude than one might reasonably expect by looking at the fluctuation of best-fit values as the underlying dataset was varied. This led to the peculiar concept of “tolerance”, namely, an a-posteriori rescaling factor of uncertainties.

Another problem is the problem of interpolation and extrapolation.

[1] The bias of an estimator (the estimator is the distribution of estimates for a particular value) is definted as the difference between the average value of the the estimator and the true value

$$\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta \tag{4.2}$$

If the bias is zero, we say that the estimator is unbiased, meaning that it gives the true value of the parameter on average.

4.1.1 Missing Higher Order Uncertainty

d for the PDF determination. Current PDF uncertainties essentially only include the propagated uncertainty arising from statistical and systematic uncertainties in the experimental data used in their determination. Methodological uncertainties related for example to the choice of functional form for the PDFs, or the fitting methodology employed, can be kept under control using closure tests, and with care can be made negligible in the data region.

Parametric uncertainties, such as those related to the value of the strong coupling $_s(m_Z)$ or the charm mass m_c can be included by performing fits for a range of parameters. However up until now MHOUs have never been included in a PDF fit: what is usually called the “PDF uncertainty” does not include the MHOU in the theoretical calculations used for PDF determination, and, more generally, does not typically include any source of theory uncertainty.

xFitter is able to perform PDF profiling and reweighting studies.

4.1.2 Fitting Procedure in xFitter

The optimal PDF values for the PDF parameters is obtained by xFitter by minimizing the chi^2 defined as

$$\chi^2 = \sum_i \frac{(\mu_i - \hat{m_i})^2}{\Delta_i^2} + \sum_\alpha b_\alpha^2 \tag{4.3}$$

Where i is the index for a given data point, μ_o its the value of the measured data point for a given observable, and $\hat{m_i}$ is the theoretical predication for data point i, and Δ_i is the uncorrelated experimental error. m_i is given by

$$\hat{m_i} = m_i + \sum_\alpha \Gamma_{i\alpha} b_\alpha \tag{4.4}$$

Where m_i is the actual theoretical value calculated using the DGLAP-evolved ODFs with given parameters $\{c_k\}$. $\Gamma_{i\alpha}$ are the correlated errors and b_α are the nuissance parameters. We see that nuissance parameters quantify the strength to correlated errors with strengths reflected by the $\Gamma_{i\alpha}$ terms. The quality of the fit can be estimated by the value of χ^2/N_{dp}, where N_{dp} is the number of data points. A value of $\chi^2/N_{dp} \approx 1$ indicates a good fit, i.e. that the agreement between the theoreticl prediction and the measured observable is at the level of the experimental uncertainties. xFitter gives two options for the choice of the χ^2 that is used in the fit: nuissance parameter representation of χ^2:

$$\chi^2(\boldsymbol{\beta}) = \sum_{i=1}^{N_\text{data}} \frac{\left(\sigma_i^\text{exp} + \sum_j \Gamma_{ij}^\text{exp} \beta_j - \sigma_i^\text{th}\right)^2}{\Delta_i^2} + \sum_j \beta_j^2 \tag{4.5}$$

or the covariance representation of χ^2

$$\chi^2(C) = \sum_{ij}^{N_\text{data}} \left(\sigma_i^\text{exp} - \sigma_i^\text{th}\right) C_{\text{tot}\,ij}^{-1} \left(\sigma_j^\text{exp} - \sigma_j^\text{th}\right) \tag{4.6}$$

The minimization algorithm used in xFitter is one implemented by the MINUIT minimization framework, called the MIGRAD algorithm. It performs a line search in the direction of the gradient and updates the covariance matrix at each step.

There are several ways to take into account the systematic and statistical correlated/uncorrelated uncertainties into the χ^2 definition in xFitter. For example, the choice for the χ^2 that is similar to the one used in the HERAPDF2.0 analysis which incorporates the combined H1 and ZEUS DIS data is

$$\chi^2(\mathbf{m},\mathbf{b}) = \sum_i \frac{\left[m_i - \Sigma_\alpha \gamma^i_\alpha \mu_i b_\alpha - \mu_i\right]^2}{\left(\delta_{i,\,\text{stat}}\sqrt{\mu_i m_i}\right)^2 + \left(\delta_{i,\,\text{uncorr}}\, m_i\right)^2} + \sum_\alpha b_\alpha^2 \tag{4.7}$$

Where $\delta_{i,\,\text{stat}}$ and $\delta_{i,\,\text{uncorr}}$ are the relative statistical and uncorrelated systematic uncertainties, respectively.

4.1.3 Parameterization: How different Parameterizations yield different PDFs

In this study we use the HERAPDF parameterization. This parameterization is is the default paramterization in xFitter,

$$xf\left(x,\mu_0^2\right) = Ax^B(1-x)^C\left[1+Dx+Ex^2\right] - A'x^{B'}(1-x)^{C'}. \tag{4.8}$$

and it is given by More specificallym the parameterization used in the HERAPDF2.0 set, at an initial scale μ_0 is

$$\begin{aligned}
xg\left(x,\mu_0^2\right) &= A_g x^{B_g}(1-x)^{C_g} - A'_g x^{B'_g}(1-x)^{C'_g} \\
xu_v\left(x,\mu_0^2\right) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left[1+E_{u_v}x^2\right] \\
xd_v\left(x,\mu_0^2\right) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}} \\
x\bar{u}\left(x,\mu_0^2\right) &= A_{\bar{u}} x^{B_{\bar{u}}}(1-x)^{C_{\bar{u}}}\left[1+D_{\bar{u}}x\right] \\
x\bar{d}\left(x,\mu_0^2\right) &= A_{\bar{d}} x^{B_{\bar{d}}}(1-x)^{C_{\bar{d}}} \\
xs\left(x,\mu_0^2\right) = x\bar{s}\left(x,\mu_0^2\right) &= r_s x\bar{d}\left(x,\mu_0^2\right) \quad r_s = \frac{f_s}{1-f_s} \quad \text{with } f_s = 0.4 \text{ fixed}
\end{aligned} \tag{4.9}$$

where the flavour basis is chosen so that

$$\begin{aligned}
u_v &= u - \bar{u} \\
d_v &= d - \bar{d}
\end{aligned} \tag{4.10}$$

We can define the light quark sea contribution as

$$S \equiv 2(\bar{u} + \bar{d}) + s + \bar{s} \tag{4.11}$$

We must remember that any PDF is subject to three constrains from number sum rules:

$$\int_0^1 \mathrm{d}x u_v \left(x, Q_0^2\right) = 2, \quad \int_0^1 \mathrm{d}x d_v \left(x, Q_0^2\right) = 1, \quad \int_0^1 \mathrm{d}x s_v \left(x, Q_0^2\right) = 0 \tag{4.12}$$

Together with the momentum sum rule

$$\int_0^1 \mathrm{d}x x \left[u_v \left(x, Q_0^2\right) + d_v \left(x, Q_0^2\right) + S \left(x, Q_0^2\right) + g \left(x, Q_0^2\right)\right] = 1 \tag{4.13}$$

There are further conditions that link the paramters. One condition is due to the quark number and momentum sum rules, that provide three constrains that fix the normalization of the gluon and valence quarks: A_g, A_{u_v} and A_{d_v}. Another condition is that the $\bar{u}$ and $\bar{d}$ are forced to behave identically at small x, giving

$$\begin{aligned} A_{\bar{u}} &= A_{\bar{d}} \\ B_{\bar{u}} &= B_{\bar{d}} \end{aligned} \tag{4.14}$$

wehre the valuence distributions are xu_v and xd_v, the gluon distribution is xg, the u-type sea $x\bar{U}$ and the d-type a

4.1.4 Comparisons of using different experimental data

In this study we use the DIS data from HERA 1+2 and Zeus combined.

4.1.5 Using xFitter

Downoad page: https://www.xfitter.org/xFitter/xFitter/DownloadPage

Manual : https://www.xfitter.org/xFitter/xFitter/DownloadPage?action=AttachFiledo=viewtarget=man

Data files: https://gitlab.cern.ch/fitters/xfitter-datafiles

Tutorials (and VM with software preinstalled): https://www.xfitter.org/xFitter/tutorials

The tutorials are from the 2016 workshop The CTEQ - MCnet School 2016

(https://indico.desy.de/event/13506/)

https://indico.desy.de/event/13506/contributions/13235/attachments/8939/10533/xfit-tutorial$_04_07_2016.pdf$

Talks:https://www.xfitter.org/xFitter/xFitter/xFitterTalks

Desy Workshop (2020) :https://indico.desy.de/event/25055/sessions/5806/20200226

Minsk workshop (2019): https://indico.desy.de/event/22011/

xFitter Krakow workshop: https://indico.desy.de/event/19213/ : Skimmed.

xFitter workshop Dubna 2016 : https://indico.cern.ch/event/458944/ : Skimmed.

xFitter Oxford workshop: https://indico.cern.ch/event/578304/ :

4.1.6 Fast Interpolation Grids

4.1.7 xFitter Tutorials

Instructions for tutorials on: https://www.slac.stanford.edu/ shoeche/mcnet16/ws/

http://www.physics.smu.edu/olness/ftp/misc2/cteq/2018/Olness/olness$_v01.pdf$

http://www.physics.smu.edu/olness/ftp/misc2/cteq/2018/Olness/notes3.txt

LOOK AT HERAFITTER TOO https://www.herafitter.org/HERAFitter/HERAFitter/HERAFitterTalks

! PDF parameterisation style. Possible styles are currently available: ! 'HERAPDF' – HERAPDF-like with uval, dval, Ubar, Dbar, glu evolved pdfs ! 'CTEQ' – CTEQ-like parameterisation ! 'CTEQHERA' – Hybrid: valence like CTEQ, rest like HERAPDF ! 'CHEB' – CHEBYSHEV parameterisation based on glu,sea, uval,dval evolved pdfs ! 'LHAPDFQ0' – use lhapdf library to define pdfs at starting scale and evolve with local qcdnum parameters ! 'LHAPDF' – use lhapdf library to define pdfs at all scales ! 'LHAPDFNATIVE'– use lhapdf library to access pdfs and alphas ! 'DDIS' – use Diffractive DIS ! 'BiLog' – bi-lognormal parametrisation

LHPDF sets available for download at: http://lhapdfsets.web.cern.ch/lhapdfsets/current/

Exercise 3 teaches how to include a new data set into an existing PDF set, without redoing a PDF fit, like Prosper paper! Although this is

to install some additional PDF set from lhapdf do

lhapdf –pdfdir=./ install CT14nlo lhapdf –pdfdir=./ install CT10nnlo lhapdf –pdfdir=./ install NNPDF30$_nlo_as_0118$

minuit.in.txt are where the **true** parameter values are defined.

]PDF studies within xFitter

4.2 xFitter

Introduction

xFitter [1] is an open-source package that provides a framework for the determination of the parton distribution functions (PDFs) of the proton for many different kinds of analyses in Quantum Chromodynamics (QCD). The xFitter project is QCD fit framework that can perform PDF fits, assess the impact of new data, compare existing PDF sets, and perform a variety of other tasks . There are a variety of options for the definition of the χ^2 function and the treatment of experimental uncertainties

xFitter is able to perform PDF profiling and reweighting studies.

4.2.1 Fitting Procedure in xFitter

The optimal PDF values for the PDF parameters is obtained by xFitter by minimizing the chi^2 defined as

$$\chi^2 = \sum_i \frac{(\mu_i - \hat{m_i})^2}{\Delta_i^2} + \sum_\alpha b_\alpha^2 \tag{4.15}$$

Where i is the index for a given data point, μ_o its the value of the measured data point for a given observable, and $\hat{m_i}$ is the theoretical predication for data point i, and Δ_i is the uncorrelated experimental error. m_i is given by

$$\hat{m_i} = m_i + \sum_\alpha \Gamma_{i\alpha} b_\alpha \tag{4.16}$$

Where m_i is the actual theoretical value calculated using the DGLAP-evolved ODFs with given parameters $\{c_k\}$. $\Gamma_{i\alpha}$ are the correlated errors and b_α are the nuissance parameters. We see that nuissance parameters quantify the strength to correlated errors with strengths reflected by the $\Gamma_{i\alpha}$ terms. The quality of the fit can be estimated by the value of χ^2/N_{dp}, where N_{dp} is the number of data points. A value of $\chi^2/N_{dp} \approx 1$ indicates a good fit, i.e. that the agreement between the theoreticl prediction and the measured observable is at the level of the experimental uncertainties. xFitter gives two options for the choice of the χ^2 that is used in the fit: nuissance parameter representation of χ^2:

$$\chi^2(\boldsymbol{\beta}) = \sum_{i=1}^{N_{\text{data}}} \frac{\left(\sigma_i^{\text{exp}} + \sum_j \Gamma_{ij}^{\text{exp}} \beta_j - \sigma_i^{\text{th}}\right)^2}{\Delta_i^2} + \sum_j \beta_j^2 \tag{4.17}$$

or the covariance representation of χ^2

$$\chi^2(C) = \sum_{ij}^{N_{\text{data}}} \left(\sigma_i^{\text{exp}} - \sigma_i^{\text{th}}\right) C_{\text{tot}\,ij}^{-1} \left(\sigma_j^{\text{exp}} - \sigma_j^{\text{th}}\right) \tag{4.18}$$

The minimization algorithm used in xFitter is one implemented by the MINUIT minimization framework, called the MIGRAD algorithm. It performs a line search in the direction of the gradient and updates the covariance matrix at each step.

There are several ways to take into account the systematic and statistical correlated/uncorrelated uncertainties into the χ^2 definition in xFitter. For example, the choice for the χ^2 that is similar to the one used in the HERAPDF2.0 analysis which incorporates the combined H1 and ZEUS DIS data is

$$\chi^2(\mathbf{m}, \mathbf{b}) = \sum_i \frac{\left[m_i - \Sigma_\alpha \gamma_\alpha^i \mu_i b_\alpha - \mu_i\right]^2}{\left(\delta_{i,\,\text{stat}} \sqrt{\mu_i m_i}\right)^2 + \left(\delta_{i,\,\text{uncorr}}\, m_i\right)^2} + \sum_\alpha b_\alpha^2 \tag{4.19}$$

Where $\delta_{i,\,\text{stat}}$ and $\delta_{i,\,\text{uncorr}}$ are the relative statistical and uncorrelated systematic uncertainties, respectively.

4.2.2 Parameterization: How different Parameterizations yield different PDFs

In this study we use the HERAPDF parameterization. This parameterization is is the default paramterization in xFitter,

$$xf\left(x, \mu_0^2\right) = Ax^B(1-x)^C\left[1 + Dx + Ex^2\right] - A'x^{B'}(1-x)^{C'}. \tag{4.20}$$

and it is given by More specificallym the parameterization used in the HERAPDF2.0 set, at an initial scale μ_0 is

$$\begin{aligned}
xg\left(x,\mu_0^2\right) &= A_g x^{B_g}(1-x)^{C_g} - A_g' x^{B_g'}(1-x)^{C_g'} \\
xu_v\left(x,\mu_0^2\right) &= A_{u_v} x^{B_{u_v}}(1-x)^{C_{u_v}}\left[1+E_{u_v}x^2\right] \\
xd_v\left(x,\mu_0^2\right) &= A_{d_v} x^{B_{d_v}}(1-x)^{C_{d_v}} \\
x\bar{u}\left(x,\mu_0^2\right) &= A_{\bar{u}} x^{B_{\bar{u}}}(1-x)^{C_{\bar{u}}}\left[1+D_{\bar{u}}x\right] \\
x\bar{d}\left(x,\mu_0^2\right) &= A_{\bar{d}} x^{B_{\bar{d}}}(1-x)^{C_{\bar{d}}} \\
xs\left(x,\mu_0^2\right) = x\bar{s}\left(x,\mu_0^2\right) &= r_s x\bar{d}\left(x,\mu_0^2\right) \quad r_s = \frac{f_s}{1-f_s} \quad \text{with } f_s = 0.4 \text{ fixed}
\end{aligned} \tag{4.21}$$

where the flavour basis is chosen so that

$$\begin{aligned}
u_v &= u - \bar{u} \\
d_v &= d - \bar{d}
\end{aligned} \tag{4.22}$$

We can define the light quark sea contribution as

$$S \equiv 2(\bar{u}+\bar{d}) + s + \bar{s} \tag{4.23}$$

We must remember that any PDF is subject to three constrains from number sum rules:

$$\int_0^1 \mathrm{d}x u_v\left(x,Q_0^2\right) = 2, \quad \int_0^1 \mathrm{d}x d_v\left(x,Q_0^2\right) = 1, \quad \int_0^1 \mathrm{d}x s_v\left(x,Q_0^2\right) = 0 \tag{4.24}$$

Together with the momentum sum rule

$$\int_0^1 \mathrm{d}x x\left[u_v\left(x,Q_0^2\right) + d_v\left(x,Q_0^2\right) + S\left(x,Q_0^2\right) + g\left(x,Q_0^2\right)\right] = 1 \tag{4.25}$$

There are further conditions that link the paramters. One condition is due to the quark number and momentum sum rules, that provide three constrains that fix the normalization of the gluon and valence quarks: A_g, A_{u_v} and A_{d_v}. Another condition is that the $\bar{u}$

and $\bar{d}$ are forced to behave identically at small x, giving

$$\begin{aligned} A_{\bar{u}} &= A_{\bar{d}} \\ B_{\bar{u}} &= B_{\bar{d}} \end{aligned} \tag{4.26}$$

wehre the valuence distributions are xu_v and xd_v, the gluon distribution is xg, the u-type sea $x\bar{U}$ and the d-type a

4.2.3 Comparisons of using different experimental data

In this study we use the DIS data from HERA 1+2 and Zeus combined.

4.2.4 Using xFitter

Downoad page: https://www.xfitter.org/xFitter/xFitter/DownloadPage

Manual : https://www.xfitter.org/xFitter/xFitter/DownloadPage?action=AttachFiledo=viewtarget=man

Data files: https://gitlab.cern.ch/fitters/xfitter-datafiles

Tutorials (and VM with software preinstalled): https://www.xfitter.org/xFitter/tutorials

The tutorials are from the 2016 workshop The CTEQ - MCnet School 2016 (https://indico.desy.de/event/13506/)

https://indico.desy.de/event/13506/contributions/13235/attachments/8939/10533/xfit-tutorial$_0 4_0 7_2 016.pdf$

Talks:https://www.xfitter.org/xFitter/xFitter/xFitterTalks

Desy Workshop (2020) :https://indico.desy.de/event/25055/sessions/5806/20200226

Minsk workshop (2019): https://indico.desy.de/event/22011/

xFitter Krakow workshop: https://indico.desy.de/event/19213/ : Skimmed.

xFitter workshop Dubna 2016 : https://indico.cern.ch/event/458944/ : Skimmed.

xFitter Oxford workshop: https://indico.cern.ch/event/578304/ :

4.2.5 Fast Interpolation Grids

4.2.6 xFitter Tutorials

Instructions for tutorials on: https://www.slac.stanford.edu/ shoeche/mcnet16/ws/

http://www.physics.smu.edu/olness/ftp/misc2/cteq/2018/Olness/olness$_v01.pdf$

http://www.physics.smu.edu/olness/ftp/misc2/cteq/2018/Olness/notes3.txt

LOOK AT HERAFITTER TOO https://www.herafitter.org/HERAFitter/HERAFitter/HERAFitterTalks

! PDF parameterisation style. Possible styles are currently available: ! 'HERAPDF' – HERAPDF-like with uval, dval, Ubar, Dbar, glu evolved pdfs ! 'CTEQ' – CTEQ-like parameterisation ! 'CTEQHERA' – Hybrid: valence like CTEQ, rest like HERAPDF ! 'CHEB' – CHEBYSHEV parameterisation based on glu,sea, uval,dval evolved pdfs ! 'LHAPDFQ0' – use lhapdf library to define pdfs at starting scale and evolve with local qcdnum parameters ! 'LHAPDF' – use lhapdf library to define pdfs at all scales ! 'LHAPDFNATIVE'– use lhapdf library to access pdfs and alphas ! 'DDIS' – use Diffractive DIS ! 'BiLog' – bi-lognormal parametrisation

LHPDF sets available for download at: http://lhapdfsets.web.cern.ch/lhapdfsets/current/

Exercise 3 teaches how to include a new data set into an existing PDF set, without redoing a PDF fit, like Prosper paper! Although this is

to install some additional PDF set from lhapdf do

lhapdf –pdfdir=./ install CT14nlo lhapdf –pdfdir=./ install CT10nnlo lhapdf –pdfdir=./ install NNPDF30$_{n}lo_{a}s_{0}118$

minuit.in.txt are where the **true** parameter values are defined.

CHAPTER 5

NNPDF Approach

5.1 PDF Determination as a Pattern Recognition Problem

The structure of the master equation 2.6 hints that the PDF could be viewed as a pattern recognition problem: given an unknown underlying function (the PDFs) that maps input instances to actually recognized outcomes (the observed cross section), use a set of data to infer the function itself.

Suppose we consider an observable, such as the cross section $\sigma_X(s, M_X^2)$ for a hard process (ie perturbatively computable in QCD) between two protons in the LHC, it as the structure

$$\sigma_X^{observed}\left(s, M_X^2\right) = \sum \int_{x_{min}}^{1} dx_1 dx_2 f_{a/A}\left(x_1, M_X^2\right) f_{b/B}\left(x_2, M_X^2\right) \sigma_{ab\to X}^{partons}\left(x_1 x_2 s, M_X^2\right) \tag{5.1}$$

Where s is the square of the center-of-mass energy (so at the LHC $s = \sqrt{14}$ TeV) and M_X is the mass of the final state; σ_X is the measurable cross section, while $\sigma_{ab\to X}^{partons}$ is the computable cross section, determined in perturbation theory from the interaction of two incoming partons a and b. $f_{a/A}$ and $f_{a/B}$ are the PDFs: they provide the information of extracting a parton of kind a, b from incoming hadrons A and B. The PDFs are a universal property of the given hadron: e.g., the proton PDFs are the same for any process with a proton in the initial state. in PDF determination one determines a probability distributions of probability distributions, i.e. a probability functional.

Correlation of uncertainties: The uncertainty on each particular PDF at a given x value, $f_i\left(x, Q_0^2\right)$ is correlated to the uncertainty on any other PDF at a different x value $f_j\left(x', Q_0^2\right)$ and this correlation must be accounted for in order to reliably estimate PDF uncertainties. Hence, PDF determination also requires the determination a covariance matrix of uncertainties in the space of probability distributions: namely, a covariance matrix functional.

The Monte Carlo representation provides a way of breaking down the problem of determining a probability in a space of functions into an (in principle infinite) set of problems in which a unique best-fit set of functions is determined. The basic idea is to turn the input probability distribution of data into a Monte Carlo representation. This means that the input data and correlated uncertainties are viewed as a probability distribution (typically, but not necessarily, a multigaussian) in the space of data, such that the central experimental values correspond to the mean and the correlated uncertainties correspond to the covariance of any two data. The Monte Carlo representation is obtained by extracting a set of replica instances from this probability distribution, in such a way that, in the limit of infinite number of replicas, the mean and and covariance over the replica sample reproduce the mean and covariance of the underlying distributions. In practice the number of replicas can be determined a posteriori by verifying that mean and covariance are reproduced to a given target accuracy.

A best-fit PDF (or rather, PDF set: i.e. one function fi(x, Q2 0) for each distinct type of parton i) is then determined for each data replica, by minimization of a suitable figure of merit. Neural networks are used to represent the PDFs, with the value of x as input, and the value of the PDF as an output (one for each PDF). [1]

5.1.1 Differences From Standard ML applications

- The PDFs are probability distributions of observables, rather than being observables themselves. In other words, PDF in determination, one determines probability distributions of probability distributions, i.e. a probability functional.
- PDF determination also requires the determination a covariance matrix of uncertainties in the space of proba- bility distributions: namely, a covariance matrix functional. between a PDF at a given x, $f_i\left(x, Q_0^2\right)$ to one at a different x' $f_j\left(x', Q_0^2\right)$

[1] **Question**: NNPDF uncertainties are within a factor of 2 from uncertainties of other PDF sets, like CTEQ, which are completely arbitrary! And each group/version has a different parameterization, how can this be? The methods are completely different yet they yield same results

5.1.2 NNPDF The Software

All the NNPDF sets are available through the LHAPDF library, see LHAPDF documentation. The names of PDF sets in LHAPDF is at https://lhapdf.hepforge.org/lhapdf5/pdfsets an example to use NNPDF sets: https://nnpdf.hepforge.org/old/html/tutorial.html The optimization in the NNPDF method consists of minimization of the χ^2

$$\chi^2 = \sum_{i,j}^{N_{\text{dat}}} (D - P)_i \sigma_{ij}^{-1} (D - P)_j \tag{5.2}$$

Where D_i is the ith data point, P_i is the product between the FastKernel tables for point i and the PDF model. σ_{ij} is the covariance matric between data points i and j, which includes both uncorrelated and correlated experimental statistical and systematic uncertainties. Multiplicative uncertainties and theory uncertainties could also be handeled aparently. **Read more on FastKernel method**

5.1.3 NN architecture

The parametrization for each PDF (or independent combination of PDFs) is

$$xf_{a/A}(x, Q_0) = A_{a/A} x^{-\alpha_{a/A}+1} (1-x)^{\beta_{a/A}} \mathrm{NN}_{a/A}(x) \tag{5.3}$$

As we see, the form of the parameterization controls the PDF behavior at small and large x, and A_i is an overall normalization constant which enforces sum rules.

5.1.4 Closure Testing by the NNPDF Collaboration

Besides validating the NNPDF methodology, closure tests also allow for uncertainty estimation in a controlled manner. Three sets of PDF data are produced: one with no experimental statistical or systematic uncertainties. A second set where by assuming the probability distribution which corresponds to the published experimental covariance matrix. A final set of data (“level 2”) is generated by taking the level 1 data as if they were actual experimen- tal data, and then applying to them the standard NNPDF methodology, which, as discussed in Section 1.2 (See Figure 2) is based on producing a set of Monte Carlo replicas of the experimental data. **They are not sure if the level 2 closure test uncertainties, i.e. the one used in the standard NNPDF procedure, are faithful**

The availability of closure test data allows performing a variety of fur- ther tests

5.1.5 Future of NNPDFs: more hyperoptimization and code-redesign, apparently

$\rightarrow$ **n3fit** $\rightarrow loweruncertainty.EverythinginthesesectionsisstandardMLproceduresandunderstand$

CHAPTER 6

Testing PDF uncertainties within xFitter and closure Testing (our approach)

6.1 Statement of the Problem

The aim of this paper is to study and quantify the uncertainties of the PDFs and their parameters, and to study whether the PDF uncertainties that are usually reported are statistically sound. As discussed below, the reported PDF uncertainties and confidence intervals make some assumptions regarding the parameterization of the PDFs, which may not be statistically justified.

From our discussion of the likelihood function in chapter , we saw that one can construct confidence intervals on the parameters using $-2\log L$: say that the likelihood for θ is multivariate normal, the likelihood function of a single observation is of the form

$$\begin{aligned} L(\boldsymbol{\theta}; x) &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}[x - g(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})]\right\} \\ &= \frac{1}{\sqrt{(2\pi)^D|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}\chi^2\right\} \end{aligned} \tag{6.1}$$

Taking the log and dropping the constant in the beginning (since the constant is independent of θ), we have

$$\begin{aligned} \log L(\boldsymbol{\theta}; \boldsymbol{x}) &= -\frac{1}{2}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}[\boldsymbol{x} - \mathrm{g}(\boldsymbol{\theta})] \\ &= -\frac{1}{2}\chi^2 \end{aligned} \tag{6.2}$$

Thus, $-2\log L$ is precisely the χ^2 expression in the leaset squares method. A popular method for estimating the errors in the maximum likelihood method is to look for parameters $\theta_\pm$ for which

$$-2\Delta \log L \equiv -2\left[\log L\left(\boldsymbol{\theta}_\pm; \boldsymbol{x}\right) - \log L(\hat{\boldsymbol{\theta}}; x)\right] = 1 \tag{6.3}$$

This method yields 68 % confidence intervals on the individual parameters. These shifts in the parameters $\boldsymbol{\theta}_{\pm}$ can be attained with the maximum likelihood methodology, as discussed in 6.1. In other words, finding the points where $\Delta\chi^2 = 1$ corresponds to the method of finding the 68 % confidence interval

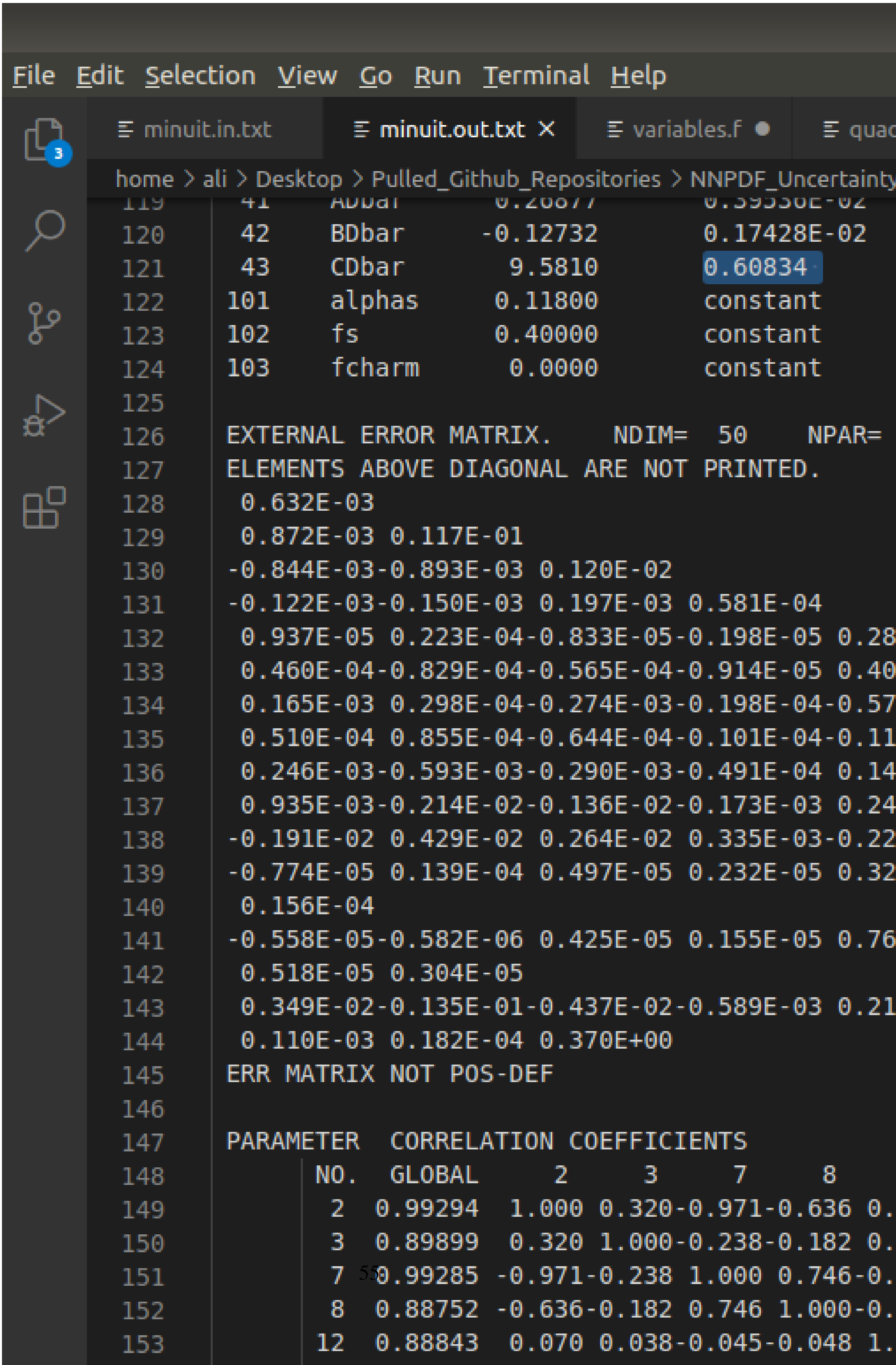
$$(\hat{\theta} - \theta_{+}, \hat{\theta} + \theta_{+}) \tag{6.4}$$

This method is widely used by all PDFs, but it is critical to remember that this assumes the normal sampling of x. If the sampling is not normal, then the probability content will be different, and must be determined for the correct sampling distribution. Since this method is used by xFitter and most other PDF fitting tools, we aim to study this assumption, whether it is true that the parameters that we aim to study are normally distributed. When xFitter gives error bars using $\chi^2 = 1$

6.2 Are the PDF parameters normally-distributed?

We start with doing a fit with xFitter using the HERAPDF parameterizations with starting values and resulting best-fit values tabulated in table **??**. The covariance matrix in the parameter space is also included in fig From here, a multivariate Gaussian sample is generated, with its mean vector being the best-fit parameter values in table $\vec{\mu} = \theta_{\text{best-fit}}$, and its covariance matrix being the covariance matrix in the parameter space.

Parameter	xFitter Name	Starting Value	Step Size	Best-Fit Value	Approximate Error
	Bg	-0.061953	0.027133	-00.61856	0.25134E-01
	Cg	5.562367	0.318464	5.5593	0.10838
	Aprig	0.166118	0.028009	0.16618	0.34574E-01
	Bprig	-0.383100	0.009784	-0.38300	0.76253E-02
	Buv	0.810476	0.016017	0.81056	0.53604E-02
	Cuv	4.823512	0.063844	4.8239	0.29342E-01
	Euv	9.921366	0.835891	9.9226	0.27481
	Bdv	1.029995	0.061123	1.0301	0.23240E-01
	Cdv	4.846279	0.295439	4.8456	0.12584
	CUbar	7.059694	0.809144	7.0603	0.22306
	DUbar	1.548098	1.096540	1.5439	0.31340
	ADbar	0.268798	0.008020	0.26877	0.39536E-02
	BDbar	-0.127297	0.003628	-0.12732	0.17428E-02
	CDbar	9.586246	1.448861	9.5810	0.60834



```
 41    ADbar      0.20877       0.39336E-02
 42    BDbar     -0.12732       0.17428E-02
 43    CDbar       9.5810       0.60834
101    alphas     0.11800       constant
102    fs         0.40000       constant
103    fcharm      0.0000       constant

EXTERNAL ERROR MATRIX.    NDIM=  50    NPAR=
ELEMENTS ABOVE DIAGONAL ARE NOT PRINTED.
 0.632E-03
 0.872E-03 0.117E-01
-0.844E-03-0.893E-03 0.120E-02
-0.122E-03-0.150E-03 0.197E-03 0.581E-04
 0.937E-05 0.223E-04-0.833E-05-0.198E-05 0.28
 0.460E-04-0.829E-04-0.565E-04-0.914E-05 0.40
 0.165E-03 0.298E-04-0.274E-03-0.198E-04-0.57
 0.510E-04 0.855E-04-0.644E-04-0.101E-04-0.11
 0.246E-03-0.593E-03-0.290E-03-0.491E-04 0.14
 0.935E-03-0.214E-02-0.136E-02-0.173E-03 0.24
-0.191E-02 0.429E-02 0.264E-02 0.335E-03-0.22
-0.774E-05 0.139E-04 0.497E-05 0.232E-05 0.32
 0.156E-04
-0.558E-05-0.582E-06 0.425E-05 0.155E-05 0.76
 0.518E-05 0.304E-05
 0.349E-02-0.135E-01-0.437E-02-0.589E-03 0.21
 0.110E-03 0.182E-04 0.370E+00
ERR MATRIX NOT POS-DEF

PARAMETER  CORRELATION COEFFICIENTS
      NO.  GLOBAL     2     3     7     8
       2  0.99294  1.000 0.320-0.971-0.636 0.
       3  0.89899  0.320 1.000-0.238-0.182 0.
       7  0.99285 -0.971-0.238 1.000 0.746-0.
       8  0.88752 -0.636-0.182 0.746 1.000-0.
      12  0.88843  0.070 0.038-0.045-0.048 1.
```

LOTS OF MORE WRITING

6.3 Determination of $\Delta\chi^2$

The confidence criterion $\Delta\chi^2$ determines the quality of the PDF fit. For example, see the Hessian method on how this is implemented. The PDF fits that employ the ideal choice $\Delta\chi^2 = 1$ are usually limited to smaller sets of data, while global fits prefer to take $\Delta\chi^2 > 1$ to account for small inconsistencies among the data sets (often from different experiments) and to compensate for the parameterization bias (the bias that is implicit in the choice of parameterization) It follows then that the corresponding uncertainty for a PDF-dependent quantity $\mathcal{O} = \mathcal{O}[f]$ can be computed as

$$(\Delta\mathcal{O})^2 = \Delta\chi^2 \sum_k \left(\frac{\partial\mathcal{O}}{\partial z_k}\right)^2 \tag{6.5}$$

In the Hessian approach, PDF error sets $S_k^\pm$ in the Hessian space are introduced

$$\begin{aligned} z\left(S_0\right) &= (0, 0, \ldots, 0), \\ z\left(S_1^\pm\right) &= \pm\sqrt{\Delta\chi^2}(1, 0, \ldots, 0) \\ z\left(S_2^\pm\right) &= \pm\sqrt{\Delta\chi^2}(0, 1, \ldots, 0) \\ &\vdots \\ z\left(S_{N_{eig}}^\pm\right) &= \pm\sqrt{\Delta\chi^2}(0, 0, \ldots, 1). \end{aligned} \tag{6.6}$$

Using these, the derivative of $\mathcal{O}$ can be calculated with a linear approximation as

$$\left(\frac{\partial\mathcal{O}}{\partial z_k}\right) \approx \frac{\mathcal{O}\left[S_k^+\right] - \mathcal{O}\left[S_k^-\right]}{2\sqrt{\Delta\chi^2}} \tag{6.7}$$

such that

$$(\Delta\mathcal{O})^2 = \frac{1}{4}\sum_k \left(\mathcal{O}\left[S_k^+\right] - \mathcal{O}\left[S_k^-\right]\right)^2 \tag{6.8}$$

6.4 Results

[pages=-,pagecommand=,width=]C$_c hap/plots.pdf$

APPENDIX A

APPENDIX

A.1 More statistics

A.2 Hypothesis Tests

The subject of hypothesis tests is very closely related to the subject of confidence intervals. A simple hypothesis is one in which the *null hypothesis* H_0 and the alternative hypothesis H_1 are completely specified. For example, if we have $H_0 : \theta = 0$ versus $H_1 : \theta = 1$, then we have a simple hypothesis, since H_0 and H_1 are both completely specified. The objective is for us to quantify when to accept or reject a hypothesis (either H_0 or H_1 since they are mutually exclusive). The way this is done is that we construct a region R in the sample space, and we say that we will reject H_0 in favor of H_1 if the observation x lies in R

$$\text{if } x \in R, \text{ Reject} H_0$$

There are two types of errors we can make

$$\text{Type I error}: \quad \text{Reject} H_0 \text{when} H_0 \text{is true}$$

$$\text{Type II error}: \quad \text{Accept} H_0 \text{when} H_1 \text{is true}$$

The probability of making a type I error

$$P(\text{Type I error}) \equiv \alpha = P(x \in R|H_0) = \int_R f(x; H_0)dx \tag{A.1}$$

The probability α is called the significance level of the test. The probability of making a type II error is

$$P(\text{Type II error}) \equiv \beta = P(x \in \bar{R}|H_0) = 1 - \int_R f(x; H_0)dx \qquad \text{(A.2)}$$

Where $\bar{R}$ is the complement of R in the sample space. Since β is the probability of incorrectly accepting H_0, $1-\beta$ is the probability of correctly rejecting H_0 (i.e. rejecting H_0 when H_1 is true, which is what we want to do). $1-\beta$ is called the power of the test.

APPENDIX B

ANOTHER SAMPLE APPENDIX

Another sample text

REFERENCES

DOCTORAL COMMITTEE

CHAIRPERSON : Dr.
Professor and Head
Department of Aerospace Engineering

GUIDE(S) : Dr. 1
Professor
Department of Aerospace Engineering

Dr. 2
Professor
Department of Aerospace Engineering

MEMBERS : Dr. A
Professor
Department of Aerospace Engineering

Dr. B
Professor
Department of Mechanical Engineering

Dr. C
Sr. Lead Research Scientist
FM Global Research, Norwood, MA, USA