Correlation and Chi2 Fitting

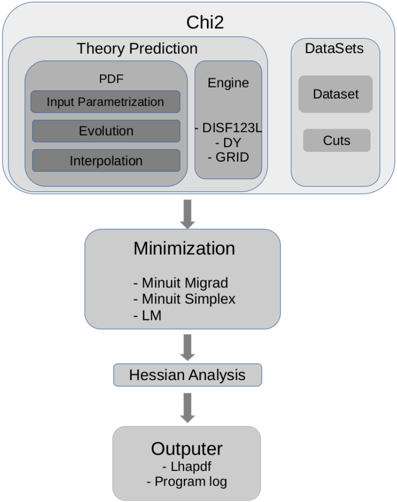
Khoirul Faiq Muzakka

WWU Münster

khoirul.muzakka@uni-muenster.de

June 18, 2020

Overview



- **Global analysis of Parton Distribution Functions** : Combine multiple datasets to constrained pdfs
- Treatment of correlated systematic errors:
  - Ignore Correlation
  - Fit nuissance parameter
  - Use Correlated Ci2

  Real case example : QCD global analysis with DIS+DY+NuTeV datasets
  - Naive χ^2 Fit : $\chi^2/N = 0.97$
  - Correlated χ^2 Fit : $\chi^2/N = 1.33$
- The interpretation of χ^2 is poorly understood in the presence of correlation between data points.
- Paradox in Parameter Error Estimation in χ^2 fitting : $\Delta\chi^2 = 1$ or $\Delta\chi^2 = \sqrt{2d}$ with d is the number of degree of freedom

Data distribution : Gaussian Errors

- Model of the data :

$$D_i = \bar{D}_i + \sigma_i r_i + \sum_\alpha s_{i\alpha} r'_\alpha \tag{1}$$

with assumption $r_i, r'_\alpha \sim \mathcal{N}(0,1)$,

$$\langle D_i \rangle = \bar{D}_i, \qquad \langle (D_i - \bar{D}_i)(D_j - \bar{D}_j) \rangle = \sigma_i^2 + \sum_\alpha s_{i\alpha} s_{j\beta} \equiv C_{ij} \tag{2}$$

The data probability distribution :

$$p(D|\bar{D}) \propto \exp\left(-\frac{1}{2}(D_i - \bar{D}_i) C_{ij}^{-1} (D_j - \bar{D}_j)\right) \tag{3}$$

- **Definition** : we say that a theory T explain the data D if there exist a theory parameter $\bar{a}$ such that

$$T_i(\bar{a}) = \bar{D}_i \tag{4}$$

for all i.

Maximum Likelihood : χ^2 as a goodness of fit measure

Under assumption that the theory $T(a)$ is correct, the data fluctuation is given by

$$D_i = T_i(a) + \sigma_i r_i + \sum_\alpha s_{i\alpha} r'_\alpha \tag{5}$$

Therefore by definition, the data distribution, given the theory parameter a, or the likelihood, is given by

$$p(D|\bar{D})\, d\mu(D) \propto \exp\left(-\frac{1}{2}(D_i - T_i(a))C_{ij}^{-1}(D_j - T_i(a))\right) \tag{6}$$

ML estimation for a is then equivalent to minimize the $\chi^2(D, a)$ function

$$\chi^2(D, a) = \sum_{i,j}(D_i - T_i(a))C_{ij}^{-1}(D_j - T_i(a)) \tag{7}$$

If all correlated errors are absent :

$$\chi^2(D, a) = \sum_i \frac{(D_i - T_i)^2}{\sigma_i^2} \tag{8}$$

χ^2 fitting : alternative goodness-of-fit

The likelihood can be obtained

$$dP(D|a) = \prod_{i,\alpha} \delta \left(D_i - T_i - \sigma_i r_i - \sum_{\gamma} \beta_{i\gamma} r'_\gamma \right) p(r_i) dr_i \; p(r'_\alpha) dr'_\alpha \, dD_i \tag{9}$$

Integrating out r_i, we obtain

$$dP(D|a) \propto \exp\left(-\frac{\tilde{\chi}^2}{2} \right) \prod_{i,\alpha} dr'_\alpha \, dD_i \tag{10}$$

$$\tilde{\chi}^2(D, a, r') = \sum_{\alpha} r'^2_\alpha + \sum_i \left(\frac{D_i - T_i - \sum_\gamma \beta_{i\gamma} r'_\gamma}{\sigma_i} \right)^2 \tag{11}$$

This form used by CTEQ collaboration for their pdf global analysis.
Pros : Easy to interpret, Can tell if the systematic errors are underestimated or overestimated.
Cons : Number of fitted parameter increase → potential problem with stability and accuracy.

Rewriting $\tilde{\chi}^2$:

$$\tilde{\chi}^2(D, a, r') = (x - A^{-1/2}B)^T(x - A^{-1/2}B) + \sum_i \left(\frac{D_i - T_i}{\sigma_i}\right)^2 - B^T A^{-1} B \tag{12}$$

with $x = A^{1/2}r'$ and

$$A_{\alpha\gamma} = \delta_{\alpha\gamma} + \sum_i = \beta_{i\alpha}\beta_{i\gamma}/\sigma_i^2, \quad B_\alpha = \sum_i \frac{\beta_{i\alpha}(D_i - T_i)}{\sigma_i^2} \tag{13}$$

$$C_{ij}^{-1} = \frac{\delta_{ij}}{\sigma_i^2} - \frac{1}{\sigma_i^2\sigma_j^2}\sum_{\alpha,\gamma}\beta_{i\alpha}(A^{-1})_{\alpha\gamma}\beta_{j\gamma} \tag{14}$$

Further integration with respect to r'_α :

$$\chi^2 = \sum_i \left(\frac{D_i - T_i}{\sigma_i}\right)^2 - B^T A^{-1} B = \sum_{i,j}(D_i - T_i)C_{ij}^{-1}(D_j - T_j) \tag{15}$$

$$= \tilde{\chi}^2(D, a, r' = B) = \min_{r'} \tilde{\chi}^2(D, a, r') \tag{16}$$

$\Longrightarrow$ **Correlated χ^2 is just the minimum of $\tilde{\chi}^2(D, a, r')$ with respect to r'.**

χ^2 fitting : Distribution of Fitted Parameters in the Linear Approximation

- **Assumption** : the theory T explain the data D, therefore there exist a parameter $\bar{a}$ such that $T(\bar{a})_i = \langle D \rangle_i$ for all i.
- Let $a^0(D)$ be the fitted parameter to the data D, namely $a^0 = \arg\min_a \chi^2(D, a)$. We assume that $a^0(D)$ is not far away from $\bar{a}$ such that the linear approximation

$$T(a^0)_i \approx \bar{T}_i + \sum_\mu \bar{T}_{i\mu}(a^0_\mu - \bar{a}_\mu) \quad (17)$$

where $\bar{T}_i = T_i(\bar{a})$ and $\bar{T}_{i\mu} = \frac{\partial T_i}{\partial a_\mu}|_{\bar{a}}$ is justified.

- The fitted parameter can be obtained as

$$\chi^2(D, a) = \sum_{i,j}(D_i - \bar{T}_i - \sum_\mu \bar{T}_{i\mu}(a_\mu - \bar{a}_\mu))C^{-1}_{ij}(D_j - \bar{T}_j - \sum_\nu \bar{T}_{j\nu}(a_\nu - \bar{a}_\nu)) \quad (18)$$

The minimum, a^0, therefore satisfy

$$a^0_\mu - \bar{a}_\mu = \sum_\nu H^{-1}_{\mu\nu}\, d_\nu, \quad H_{\mu\nu} = \sum_{i,j} \bar{T}_{i\mu} C^{-1}_{ij} \bar{T}_{j\nu} \quad (19)$$

$$d_\mu = \sum_{i,j} \bar{T}_{i\mu} C^{-1}_{ij}(D_i - \bar{T}_i) \to \langle d_\mu \rangle = 0 \quad (20)$$

which implies $\langle a^0(D)_\mu \rangle = \bar{a}_\mu$ and $\langle (a^0_\mu - \bar{a}_\mu)(a^0_\nu - \bar{a}_\nu) \rangle = H^{-1}_{\mu\nu}$

χ^2 fitting : Distribution of Fitted Parameters in the Linear Approximation

- The fitted parameters then are normally distributed around $\bar{a}$ with correlation $H^{-1}_{\mu\nu}$.

$$p(a^0) \propto \exp\left(-\frac{1}{2}\sum_{\mu,\nu}(a^0_\mu - \bar{a}_\mu)H_{\mu\nu}(a^0_\mu - \bar{a}_\mu)\right) \tag{21}$$

- The value of χ^2 at minimum is given by

$$\chi^2(D, a^0) = \sum_{i,j}(D_i - \bar{T}_i)C^{-1}_{ij}(D_j - \bar{T}_j) - \sum_{\mu,\nu}(a^0_\mu - \bar{a}_\nu)H_{\mu\nu}(a^0_\nu - \bar{a}_\nu) \tag{22}$$

which implies

$$\langle\chi^2(D, a^0)\rangle = N - d, \quad \left\langle\left(\chi^2(a^0) - \langle\chi^2(a^0)\rangle\right)^2\right\rangle = 2(N - d) \tag{23}$$

- For any a close to $a^0(D)$, one can prove that

$$\chi^2(D, a) = \chi^2(D, a^0) + \sum_{\mu,\nu}(a - a^0)_\mu H_{\mu\nu}(a - a^0)_\nu \tag{24}$$

Therefore, we have validated the following identification

$$H_{\mu\nu} = \frac{1}{2}\left.\frac{\partial^2\chi^2(D, a)}{\partial a_\mu \partial a_\nu}\right|_{a^0} \tag{25}$$

Note that, although the hessian seemingly depend on D, it is actually independent of it.

χ^2 fitting : Errors of observables

- The variance of the fitted parameters $\langle (a^0_\mu - \bar{a}_\mu)(a^0_\nu - \bar{a}_\nu) \rangle = H^{-1}_{\mu\nu}$, hence the uncertainty of the fitted parameters is given by

$$\delta a^0_\mu = \sqrt{H^{-1}_{\mu\mu}} \tag{26}$$

- Defining a new set of parameters

$$z^0(D) = \hat{H}^{1/2} U^T (a^0 - \bar{a}) \tag{27}$$

where $H = U \hat{H} U^T$. Due to data fluctuation, z^0 is distributed according to $z^0 \sim \mathcal{N}(0, 1)$ with

$$\langle z^0_\mu \rangle = 0, \qquad \langle z^0_\mu z^0_\nu \rangle = \delta_{\mu\nu} \tag{28}$$

In terms of z^0, the χ^2 at minimum is given by

$$\chi^2(a^0, D) = \chi^2(\bar{a}, D) + {z^0}^T z^0 \tag{29}$$

Thus, $z^0_\mu z^0_\nu = \delta_{\mu\nu}$ implies $|\chi^2(a^0, D) - \chi^2(\bar{a}, D)| = 1$.

- Defining $z = \hat{H}^{1/2} U^T (a - a^0)$, the log-likelihood at some point a close to a^0 is given by

$$\chi^2(D, a) = \chi^2(D, a^0) + z^T z \tag{30}$$

therefore, 1σ deviation from $\bar{a}$ correspond to $\Delta\chi^2 = 1$.

- Lets assume instead of $\Delta\chi^2 = 1$ we set $\Delta\chi^2 = T^2$ for some T.
- For any observable $X(a(z))$, the displacement of X due to displacement in z around minimum $z = 0$ is given by

$$\Delta X(a) = \sum_\mu \left.\frac{\partial X}{\partial z_\mu}\right|_{z=0} z_\mu = \nabla_z X.\vec{z} \tag{31}$$

- Given the tolerance $\Delta\chi^2 \equiv T^2$, the displacement z consistent with $z^T z$ can be taken as

$$\vec{z} = \frac{\nabla_z X}{|\nabla_z X|} T \to \Delta X(a) \leq T\,|\nabla_z X| \tag{32}$$

- Define vectors in the z-space

$$(\vec{z}^{\mu}_{\pm})_\nu = \pm T\,\delta_{\mu\nu} \tag{33}$$

- The derivative can be estimated using finite difference

$$\left.\frac{\partial X}{\partial z_\mu}\right|_{z=0} = \frac{X(z^\mu_+) - X(z^\mu_-)}{2T} \tag{34}$$

- The error of X due to uncertainties of fitted parameters is therefore given by

$$\delta X = \frac{1}{2}\sqrt{\sum_\mu \left[X(z^\mu_+) - X(z^\mu_-)\right]^2} \tag{35}$$

- As an application, we can use this formula to calculate the uncertainties of the fitted parameters. We know that $a(z) = U\hat{H}^{1/2}z$, hence

$$a_\mu(z_+^\nu) - a_\mu(z_+^\nu) = 2T(U\hat{H}^{-1/2})_{\mu\nu} \tag{36}$$

$$\Rightarrow (\delta a_\mu)^2 = \frac{1}{2}\sqrt{4T^2(U\hat{H}^{-1}U^T)_{\mu\mu}} = T\sqrt{H^{-1}_{\mu\mu}} \tag{37}$$

which for $T = 1$ agrees with (26).

Simulations

See notebook

Conclusions

- Do not ignore data correlation during χ^2 fitting.
- Correlated χ^2 can be interpreted as a least square loss function with the residual is given by $D_i - T_i(a) - \sum_\alpha \beta_{i\alpha} r'_\alpha$ with $r'_\alpha(a, D)$ is the fitted systematic fluctuation.
- The tolerance $\Delta\chi^2 = 1$ is valid theoretically under some reasonable assumptions if the theory describe the data well.

References

Kovarik, Nadolsky, Soper (2012)
Hadron Structure in High Energy Collisions

Eadie, Drijard, James, Roos, Sadoulet (2006)
Statistical Method in Experimental Physics