Bayesian perspective on QCD global analysis

Nobuo Sato
University of Connecticut/JLab
DIS18,
Kobe, Japan,
April 16-20, 2018

In collaboration with:
- A. Accardi
- E. Nocera
- W. Melnitchouk
Bayesian methodology in a nutshell

- In QCD global analysis PDFs are parametrized at some scale $Q_0$. e.g.

$$f(x) = N x^a (1 - x)^b (1 + c \sqrt{x} + dx + \ldots)$$

$$f(x) = N x^a (1 - x)^b \text{NN}(x; \{\theta, w_i\})$$

- “fitting” is essentially estimation of

$$E[f] = \int d^n a \; \mathcal{P}(a|\text{data}) \; f(a)$$

$$V[f] = \int d^n a \; \mathcal{P}(a|\text{data}) \; (f(a) - E[f])^2$$

- The probability density $\mathcal{P}$ is given by the Bayes’ theorem

$$\mathcal{P}(f|\text{data}) = \frac{1}{Z} \mathcal{L}(\text{data}|f) \pi(f)$$
Bayesian methodology in a nutshell

- The likelihood function is not unique. A standard choice is the Gaussian likelihood

\[ \mathcal{L}(d|a) = \exp \left[ -\frac{1}{2} \sum_i \left( \frac{d_i - \text{th} y_i(a)}{\delta d_i} \right)^2 \right] \]

- Priors are designed to veto unphysical regions in parameter space. E.g.

\[ \pi(a) = \prod_i \theta(a_i - a_{i\text{\tiny min}})\theta(a_{i\text{\tiny max}} - a_i) \]

- How do we compute \( E[f], V[f] \)?
  - Maximum likelihood
  - Monte Carlo methods
Maximum Likelihood

- Estimation of expectation value

\[
E[f] = \int d^m a \ P(a|data) \ f(a) \approx f(a_0)
\]

- \(a_0\) is estimated from optimization algorithm

\[
\max[P(a|data)] = P(a_0|data)
\]

\[
\max[L(data|a)\pi(a)] = L(data|a_0)\pi(a_0)
\]

- or equivalently Chi-squared minimization

\[
\min[-2 \log(L(data|a)\pi(a))] = -2 \log(L(data|a_0)\pi(a_0)) = \sum \left( \frac{d_i - thy_i(a_0)}{\delta d_i} \right)^2 - 2 \log(\pi(a_0)) = \chi^2(a_0) - 2 \log(\pi(a_0))
\]
Maximum Likelihood

- Estimation of variance (Hessian method)

\[
V[f] = \int d^n a \ P(a|data) \ (f(a) - E[f])^2
\]
\[
\approx \sum_k \left( \frac{f(t_k = 1) - f(t_k = -1)}{2} \right)^2
\]

- It relies on factorization of \( P(a|data) \) along eigen directions

\[
P(a|data) \propto \prod_k \exp \left( -\frac{1}{2} t_k^2 \right) + O(\Delta a^3)
\]

- and linear approximation of \( f(a) \)

\[
(f(a) - E[f])^2 = \left( \sum_k \frac{\partial f}{\partial t_k} t_k \right)^2 + O(a^3)
\]
Maximum Likelihood

- **pros**
  + Very practical. Most PDF groups use this method
  + It is computationally inexpensive
  + $f$ and its eigen directions can be precalculated/tabulated

- **cons**
  + Assumes local Gaussian approximation of the likelihood
  + Assumes linear approximation of the observables $\mathcal{O}$ around $a_0$
  + The assumptions are strictly valid for linear models.
  + Computation of the Hessian matrix is numerically unstable if flat directions are present

- **examples**
  
  $\rightarrow$ if $f(x) = a + bx + cx^2$ then $E[f(x)] = E[a] + E[b]x + E[c]x^2$
  
  $\rightarrow$ but $f(x) = N x^a (1 - x)^b$ then $E[f(x)] \neq E[N] x^{E[a]} (1 - x)^{E[b]}$
Monte Carlo Methods

- Recall that we are interested in computing

\[
E[f] = \int d^n a \ P(a|data) \ f(a)
\]

\[
V[f] = \int d^n a \ P(a|data) \ (f(a) - E[f])^2
\]

- Any MC method attempts to do this using MC sampling

\[
E[f] \simeq \sum_k w_k f(a_k)
\]

\[
V[f] \simeq \sum_k w_k (f(a_k) - E[f])^2
\]

- i.e to construct the sample distribution \( \{w_k, a_k\} \) of the parent distribution \( P(a|data) \)
Monte Carlo Methods

- Resampling + cross validation
- Nested Sampling (NS)
- Hybrid Markov chain (HMC); Gabin Gbedo, Mangin-Brinet (2017)

Nested Sampling, Lin et al (2018)

Resampling + cross validation (R+CV)

- Resample the data points within quoted uncertainties using Gaussian statistics

\[ d_{k,i}^{\text{(pseudo)}} = d_i^{\text{(exp)}} + \sigma_i^{\text{(exp)}} R_{k,i} \]

- Fit each pseudo data sample \( k = 1, \ldots, N \) to obtain parameter vectors \( a_k \):

\[ P(a|\text{data}) \rightarrow \{ w_k = 1/N, a_k \} \]

- For large number of parameters, split the data into training and validation sets and find \( a_k \) that best describes the validation sample
**Nested Sampling (NS)**

- **The basic idea**: compute

\[
Z = \int \mathcal{L}(\text{data} | \alpha) \pi(\alpha) d^n\alpha = \int_0^1 \mathcal{L}(X) dX
\]

+ The procedure collects samples from isolikelihoods and they are weighted by their likelihood values
+ Insensitive to local minima → faithful conversion of

\[
\mathcal{P}(\alpha | \text{data}) \rightarrow \{w_k, \alpha_k\}
\]

+ Multiple runs can be combined into one single run → the procedure can be parallelized

---

- arxiv.org/abs/1703.09701
Comparison between the methods

- Given a likelihood, does the evaluation of $E[f]$ and $V[f]$ depend on the method? → use stress testing numerical example

- Setup:
  - Simulate a synthetic data via rejection sampling
  - Estimate $E[f]$ and $V[f]$ using different methods
Comparison between the methods

- HESS, NS and R provide the same uncertainty
- R+CV over estimates the uncertainty by roughly a factor of 2
- Uncertainties also depends on training fraction (tf)
- The results confirmed also within a neural net parametrization
Beyond gaussian likelihood

- The Gaussian likelihoods are not adequate to describe uncertainties in the presence of incompatible data sets
- Example:
  - Two measurements of a quantity \( m \):
    \((m_1, \delta m_1), (m_2, \delta m_2)\)
  - The expectation value and variance can be computed exactly

\[
E[m] = \frac{m_1 \delta m_2 + m_2 \delta m_1}{\delta m_2^2 + \delta m_1^2}
\]

\[
V[m] = \frac{\delta m_2^2 \delta m_1^2}{\delta m_2^2 + \delta m_1^2}
\]

- Note: \( V[m] \) is independent of \(|m_1 - m_2|\)

- To obtain more realistic uncertainties, the likelihood function needs to be modified. (e.g. Tolerance criterion)
Likelihood profile in CJ15

- 24 parameters, 33 data sets
- Eigen direction without incompatibilities
24 parameters, 33 data sets

Eigen direction with incompatibilities

Modified likelihood function is needed
Beyond gaussian likelihood

- Tolerance criterion (standard choice)
- Disjoint likelihood function. e.g.

**joint:**

\[ \mathcal{L}(m_1, m_2 | m; \delta m_1 \delta m_2) = \mathcal{L}(m_1 | m; \delta m_1) \mathcal{L}(m_2 | m; \delta m_2) \]

\[ E[m] = \frac{m_1 \delta m_2 + m_2 \delta m_1}{\delta m_2^2 + \delta m_1^2} \quad V[m] = \frac{\delta m_2^2 \delta m_1^2}{\delta m_2^2 + \delta m_1^2} \]

**disjoint:**

\[ \mathcal{L}(m_1, m_2 | m; \delta m_1 \delta m_2) = \frac{1}{2} (\mathcal{L}(m_1 | m; \delta m_1) + \mathcal{L}(m_2 | m; \delta m_2)) \]

\[ E[m] = \frac{1}{2} (m_1 + m_2) \quad V[m] = \frac{1}{2} (\delta m_1^2 + \delta m_2^2) + \left( \frac{m_1 - m_2}{2} \right)^2 \]

- Empirical Bayes, hierarchical Bayes ...
- Many alternatives still to be explored
Summary and outlook

- Bayesian formulation for global analysis provides a more general perspective for global fits than the traditional chi-squared minimization.

- MC approaches are useful to explore new likelihood functions and priors.

- Uncertainties on PDFs depend on parametrization as well as assumptions about the likelihood function and the priors.

- Given the likelihood function and priors, uncertainties on PDFs should be independent of the parametrization in the region where PDFs can be constrained.

- Also the results should be independent of the MC sampling method.