A statistical description of theory uncertainty from missing higher orders

Marco Bonvini

INFN, Rome 1 unit

DIS2019, 11 April 2019
How accurate are current theoretical predictions?

WHAT PRECISION AT NNLO?

For many processes NNLO scale band is $\sim \pm 2\%$

Though only in 3/17 cases is NNLO (central) within NLO scale band…
How do we determine theory uncertainties?

**Scale variation:** dependence on unphysical scales of a physical observable at $N^n$LO is of higher order

$$
\Sigma_{N^n\text{LO}}(\mu) = \sum_{k=0}^{n} c_k(\mu) \alpha_s^k(\mu) \quad \mu \frac{d}{d\mu} \Sigma_{N^n\text{LO}}(\mu) = \mathcal{O}(\alpha_s^{n+1})
$$

Canonical uncertainty: variation by a factor of 2 about a “central” scale $\mu_0$

$$
\Sigma_{\text{true}} \approx \Sigma_{N^n\text{LO}}(\mu_0) \pm \max_{\mu_0/2 \leq \mu \leq 2\mu_0} |\Sigma_{N^n\text{LO}}(\mu) - \Sigma_{N^n\text{LO}}(\mu_0)|
$$

Other (less used) methods:

- use the last known order: $\Sigma_{\text{true}} \approx \Sigma_{N^n\text{LO}} \pm |\Sigma_{N^n\text{LO}} - \Sigma_{N^{n-1}\text{LO}}|

- include other scales or parameters to get more variations (e.g. in resummed calculations)

- novel approach recently proposed by F. Tackmann [slides at SCET2019]

- ....
Caveats of canonical methods

\[ \Sigma_{\text{true}} \approx \Sigma_{N^{n\text{LO}}(\mu_0)} \pm \max_{\mu_0/2 \leq \mu \leq 2\mu_0} \left| \Sigma_{N^{n\text{LO}}(\mu)} - \Sigma_{N^{n\text{LO}}(\mu_0)} \right| \]

**Which central scale \( \mu_0 \)?**
There are guiding principles, but several values are acceptable

**How much should I vary the scale?**
The factor of 2 is an arbitrary choice

**How do I interpret the uncertainty?**
I can say that it’s a range where I’m sure the next order will lie, or the 1-sigma of a gaussian, but it’s very arbitrary...

---

**Need for a statistically-sound definition of theoretical uncertainties, which does not depend so much on arbitrary assumptions**

Ideally, theory uncertainty from missing higher orders should be a probability distribution

This would also allow for a statistically meaningful comparison of theory predictions with data (e.g. in a PDF fit, or in precision tests of the SM)
Cacciari and Houdeau [1105.5152] proposed a statistical model for the interpretation of theory uncertainties, from which one can compute the probability distribution for the observable $\Sigma$ given the first few terms

$$
\Sigma = \sum_k c_k \alpha_s^k
$$

“*We make the assumption that all the coefficients $c_k$ in a perturbative series share some sort of upper bound $\bar{c} > 0$ to their absolute values, specific to the physical process studied. The calculated coefficients will give an estimate of this $\bar{c}$, restricting the possible values for the unknown $c_k$.*”

In other words, the model assumes that

$$
|c_k| \leq \bar{c} \quad \forall k
$$

implying that the perturbative expansion is bounded by a geometric series

$$
|\Sigma| = \left| \sum_k c_k \alpha_s^k \right| \leq \sum_k \bar{c} \alpha_s^k = \frac{\bar{c}}{1 - \alpha_s}
$$
Caveats of the Cacciari-Houdeau model

- it assumes a convergent perturbative expansion
  - acceptable from an asymptotic expansion point of view
  - attempt to treat the series as factorially divergent
    - [Bagnaschi, Cacciari, Guffanti, Jenniches 1409.5036]

- if the coefficients grow as a power, \( c_k \sim \eta^k \), which is very likely, the method cannot perform well
  - Cacciari-Houdeau proposed a modified version with \( \eta \) accounted for
  - in [Bagnaschi, Cacciari, Guffanti, Jenniches 1409.5036] \( \eta \) is determined from a survey on various observables
  - in an alternative approach [Forte, Isgrò, Vita 1312.6688] the value of \( \eta \) is fitted

- it still depends on the choice of the “central” scale
In this talk

\[ \Sigma = \Sigma_{\text{LO}}(\mu) \sum_{k \geq 0} \delta_k(\mu) \quad \Sigma_{\text{LO}}(\mu) \delta_k(\mu) = c_k(\mu) \alpha_s^k(\mu) \]

I propose two new methods based on the following assumptions:
In this talk

\[ \Sigma = \Sigma_{LO}(\mu) \sum_{k \geq 0} \delta_k(\mu) \]

\[ \Sigma_{LO}(\mu)\delta_k(\mu) = c_k(\mu)\alpha_s^k(\mu) \]

I propose two new methods based on the following assumptions:

- **geometric behaviour model (i.e., improved Cacciari-Houdeau)**

\[ |\delta_k(\mu)| \leq c a^k \quad \forall k \]

depends on two hidden parameters \( c, a \), and also on \( \mu \)
In this talk

\[ \Sigma = \Sigma_{LO}(\mu) \sum_{k \geq 0} \delta_k(\mu) \]

\[ \Sigma_{LO}(\mu) \delta_k(\mu) = c_k(\mu) \alpha_s^k(\mu) \]

I propose two new methods based on the following assumptions:

- **geometric behaviour model (i.e., improved Cacciari-Houdeau)**
  
  \[ |\delta_k(\mu)| \leq ca^k \quad \forall k \]

  depends on two hidden parameters \( c, a \), and also on \( \mu \)

- **scale variation inspired model**

  \[ |\delta_k(\mu)| \leq \lambda r_{k-1}(\mu) \quad \forall k, \quad r_k(\mu) \approx \left| \mu \frac{d}{d\mu} \log \Sigma_{N^k_{LO}}(\mu) \right| \]

  depends on one hidden parameter \( \lambda \), and also on \( \mu \)

- Assumptions not contradictory: we can assume both at the same time!
Model 1: Geometric behaviour

The condition

$$|\delta_k(\mu)| \leq ca^k$$

is implemented through the likelihood

$$P(\delta_k|c, a, \mu) \propto \theta(ca^k - |\delta_k(\mu)|) = ca^k - ca^k\delta_k \lambda_{r_k-1}\delta_k - \lambda_{r_k-1}cak-cak \delta_k$$

namely all values of $\delta_k$ within the allowed range are equally likely

Inference scheme:

$$\begin{align*}
\delta_0, \ldots, \delta_n & \quad \text{inference} & c, a & \quad \text{inference} & \delta_{n+1}, \delta_{n+2}, \ldots & \quad \text{sum} & \Sigma \\
\text{known} & & & & \text{unknown} & & 
\end{align*}$$

Final output:

$$P(\Sigma|\tilde{\delta}_0, \ldots, \tilde{\delta}_n, \mu, \text{model}_1)$$

Note: The assumption is good for convergent series, but a power divergence is also possible if $a > 0$ is allowed
Model 2: Scale variation inspired

The condition

$$|\delta_k(\mu)| \leq \lambda r_{k-1}(\mu)$$

is implemented through the likelihood

$$P(\delta_k | r_{k-1}, \lambda, \mu) \propto \theta(\lambda r_{k-1} - |\delta_k(\mu)|) = \begin{cases} c_a k & \text{if } \lambda_{r_{k-1}} \leq |\delta_k(\mu)| \\ c_a k - \lambda_{r_{k-1}} \delta_k & \text{if } \lambda_{r_{k-1}} > |\delta_k(\mu)| \end{cases}$$

namely all values of $\delta_k$ within the allowed range are equally likely

Inference scheme:

$$\begin{align*}
\delta_0, \ldots, \delta_n, r_0, \ldots, r_{n-1} & \xrightarrow{\text{inference}} \lambda \\
\delta_0, \ldots, \delta_n, r_0, \ldots, r_{n-1} & \xrightarrow{\text{inference+}r_n} \delta_{n+1} \\
\delta_0, \ldots, \delta_n, \mu, \text{model}_2 & \xrightarrow{\text{sum}} \Sigma_{N^{n+1}LO}
\end{align*}$$

in this case only the first missing higher order can be predicted:

$$P(\Sigma_{N^{n+1}LO} | \delta_0, \ldots, \delta_n, \mu, \text{model}_2)$$

Note: Canonical scale variation is $\sim$ recovered for $\lambda = \log 2$; having $\lambda$ a free parameter generalizes canonical scale variation to any size of the variation.
Construct a scale-independent result

Basic idea: treat the unphysical scale $\mu$ as a parameter of the model, and simply marginalize it

$$P(\Sigma|\delta_0, ..., \delta_n) = \int d\mu \ P(\Sigma|\delta_0, ..., \delta_n, \mu)P(\mu|\delta_0, ..., \delta_n)$$

where $P(\mu|\delta_0, ..., \delta_n)$ is the posterior distribution for $\mu$ given the known orders

This distribution depends on the model

In this approach, inference on $\mu$ selects the values that give the best convergence properties according to the model

The prior $P_0(\mu)$ contains our prejudices on what are the most appropriate scales, but the results are largely independent on the precise form and size of the prior $\Rightarrow$ a lot of arbitrariness is removed!
Some examples of observables known at high perturbative order in QCD:

- Higgs production in gluon fusion ($ggH$) inclusive cross section ($N^3\text{LO}$) (converges slowly)
- $e^+e^- \rightarrow \text{hadrons} (N^4\text{LO})$ (converges fast)
- ....

Some applications to series with known sums to verify the goodness of the model:

- purely resummed $ggH$ at $N^3\text{LL}$, expanded in powers of $\alpha_s$
- factorially divergent series with alternating signs $\sum_k (-1)^k k!\alpha_s^k$
- factorially divergent series with equal signs $\sum_k k!\alpha_s^k$
- ....
Higgs in gluon fusion at LHC: the “raw” result

\[ m_H = 125 \text{ GeV at LHC 13 TeV in the rEFT} \]
Higgs in gluon fusion at LHC: models at work

$\sigma = 125$ GeV at LHC 13 TeV in the rEFT

$\mu_F = m_H/2$

$\mu_R = m_H/2$

Probability distribution of the cross section: $P(\sigma)$

$P(\Sigma_{NLO} | \delta_0, \mu = m_H/2)$
Higgs in gluon fusion at LHC: models at work

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

\[ \mu_R = \frac{m_H}{2} \]

\[ P(\Sigma_{\text{NNLO}}|\delta_0, \delta_1, \mu = \frac{m_H}{2}) \]
$m_H = 125$ GeV at LHC 13 TeV in the rEFT

Probability distribution of the cross section:

$P(\sigma)\quad\sigma [\text{pb}]$

- NLO given LO
- N2LO given LO,NLO
- N3LO given LO,NLO,N2LO

solid: geometric behaviour model
dashed: scale variation model

$\mu_F = m_H/2$
$\mu_R = m_H/2$

$P(\Sigma_{N^3LO}; \delta_0, \delta_1, \delta_2, \mu = m_H/2)$
Higgs in gluon fusion at LHC: models at work

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

Probability distribution of the cross section: $P(\sigma)$

solid: geometric behaviour model
dashed: scale variation model

$\mu_R = m_H/2$

$P(\sum_{N=0}^{N=4} |\delta_0, \delta_1, \delta_2, \delta_3, \mu = m_H/2)$
Higgs in gluon fusion at LHC: models at work

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

Probability distribution of the cross section: $P(\sigma)$

- NLO given LO
- N2LO given LO, NLO
- N3LO given LO, NLO, N2LO
- N4LO given LO, NLO, N2LO, N3LO

solid: geometric behaviour model

dashed: scale variation model

$\mu_R = m_H$

$P(\Sigma^{N4\text{LO}}|\delta_0, \delta_1, \delta_2, \delta_3, \mu = m_H)$
Higgs in gluon fusion at LHC: models at work

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

Scale-independent result

$$P(\Sigma_{N^4LO} | \delta_0, \delta_1, \delta_2, \delta_3)$$
Higgs in gluon fusion at LHC: final results

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

conventional result: canonical scale variation by a factor of 2 about $\mu_R = m_H/2$ (best convergence properties)

new result: scale variation inspired model

Marco Bonvini
Higgs in gluon fusion at LHC: final results

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

$\mu_F = m_H/2$

$m_H = 125$ GeV

conventional result: canonical scale variation by a factor of 2 about $\mu_R = m_H/2$ (best convergence properties)

new result: geometric behaviour model

\[ \sigma \text{ [pb]} \]

knowledge of

LO
NLO
NNLO
$N^3$LO

std dev
95\% DoB
68\% DoB
median
mode
mean
conventional result:
Validation using known sums

purely resummed $N^3\text{LL}$ cross section for $ggH$

$\sigma = 125 \text{ GeV}$
$\mu_F = m_H/2$

new results:
left: scale variation inspired model
right: geometric behaviour model

Factorially divergent series with alternating sign

Factorially divergent series with same sign

Marco Bonvini
A statistical description of theory uncertainty from missing higher orders
Two new statistical models for theory uncertainties:
- an improved version of Cacciari-Houdeau (geometric behaviour model)
- a model inspired by scale variation

A novel way to obtain \textit{scale-independent results}

Preliminary results very promising (geom. behaviour model somewhat better)

What’s next?
- A systematic study of priors
- Investigate potential improvements/combinations of the models
- Release a \textbf{public code}
- Consider \textbf{correlations} among different kinematic configurations
- Applications
Backup slides
$e^+ e^- \rightarrow \text{hadrons: the “raw” result}$

Ratio $e^+ e^- \rightarrow \text{hadrons}$ for $Q = 10$ GeV normalized to 1 at LO

- $\sigma [\text{pb}]$ vs $\mu_R/m_H$

- LO
- NLO
- NNLO
- N3LO
- N4LO

Marco Bonvini

A statistical description of theory uncertainty from missing higher orders
$e^+e^- \rightarrow \text{hadrons: final results}$

Ratio $e^+e^-\rightarrow$hadrons for $Q = 10$ GeV normalized to LO

conventional result: canonical scale variation by a factor of 2 about $\mu_R = Q$

new results:
left: scale variation model
right: geometric behaviour model

Marco Bonvini
A statistical description of theory uncertainty from missing higher orders
Ratio $e^+e^->\text{hadrons}$ for $Q = 10$ GeV normalized to LO

conventional result: canonical scale variation by a factor of 2 about $\mu_R = Q$

new results:
left: scale variation model
right: geometric behaviour model
The priors

**geometric behaviour model**

\[
P_0(c, a) = P_0(c)P_0(a)
\]

\[
P_0(c) = \frac{\epsilon}{c^{1+\epsilon}}\theta(c - 1) \quad \epsilon = 0.1
\]

\[
P_0(a) = 2(1 - a)\theta(a)\theta(1 - a)
\]

**scale variation inspired model**

\[
P_0(\lambda) = \lambda e^{-\lambda}
\]

NB: These distributions assume the trivial knowledge of \(\delta_0 = 1\). The first non-trivial information updating these distributions is \(\delta_1\)
Posterior of the geometric-behaviour model parameters

$ggH:$

$e^+e^- \rightarrow \text{hadrons}:$
Posterior of the scale variation inspired model parameter $ggH$:

$e^+e^- \rightarrow \text{hadrons}$:
Posterior of $\mu$ in $ggH$

Geometric-behaviour model

$m_H = 125$ GeV at LHC 13 TeV in the rEFT

$\mu_F = m_H/2$

Support: $m_H/10 < \mu < 2m_H$