Theory Uncertainties.

(aka The Ugly)

Frank Tackmann

Deutsches Elektronen-Synchrotron

BIRS workshop on Systematic Effects and Nuisance Parameters in Particle Physics Data Analyses
April 25, 2023

European Research Council
Established by the European Commission

Disclaimers and Apologies.

- I'm not an experimentalist let alone a statistics expert, so apologies if some things are too pedestrian and others too complicated ...
- I have tried to abstract things out as much as I could, but please interupt if I slip too much into theory slang
  - I'm also more than happy to go into more detail ...
- There are many opinions about theory uncertainties (usually as many as there are theorists in the room ...)
  - So in matters opinion I will give you mine ...

What Are We Talking About?

Pendulum example

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{g}{l}\sin\theta = 0 \qquad \overset{\theta\ll 1}{\Longrightarrow} \qquad \theta(t) = \theta_0 \cos\frac{2\pi t}{T}\,, \quad T = 2\pi\sqrt{\frac{l}{g}}$$

- We have a formula to obtain the quantity of interest (g) from our measurement ($\theta(t)$ or T)
- This formula is the *theory prediction*
- The *theory uncertainty* is due to the fact that the formula itself is not exact but derived in some approximation ($\theta \ll 1$)
  - It is *not* the inexact knowledge of parameters needed in the (otherwise exact) formula (e.g. the length l of the pendulum)

    These are the usual systematics (parametric uncertainties)
  - Note: Sometimes certain parametric uncertainties are also called a theory uncertainty just because they primarily enter via the theory predictions (e.g. parton distribution functions).

    For this talk these are not theory uncertainties.

What Are We Talking About?

Pendulum example

$$\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2} + \frac{g}{l}\sin\theta = 0 \qquad \overset{\theta \ll 1}{\Longrightarrow} \qquad \theta(t) = \theta_0 \cos\frac{2\pi t}{T}\,, \quad T = 2\pi\sqrt{\frac{l}{g}}$$

- We have a formula to obtain the quantity of interest (g) from our measurement ($\theta(t)$ or T)
- This formula is the *theory prediction*
- The *theory uncertainty* is due to the fact that the formula itself is not exact but derived in some approximation ($\theta \ll 1$)

$\Rightarrow$ The Challenge: How to account for the inexactness of the formula itself?

- The theory uncertainty is different from other systematics because a priori there is no auxiliary measurement to improve inexactness
- But wait until the end of the talk ...

In Particle Physics.

Data $\longrightarrow$ Measured cross sections (or limits) $\longrightarrow$ Lagrangian parameters

Theory predictions $\uparrow$ · *Theory predictions* $\uparrow$

- In one way or another, we always compare a measured quantity to its theory prediction

$$f^{\text{measured}} = f^{\text{predicted}}(p_i)$$

  - where p_i are the parameter(s) of interest to be determined
  - Exactly how and where this comparison happens is not relevant for now

In Particle Physics.

Data $\longrightarrow$ Measured cross sections (or limits) $\longrightarrow$ Lagrangian parameters

Theory predictions   *Theory predictions*

- In one way or another, we always compare a measured quantity to its theory prediction

$$f^{\text{measured}} = f^{\text{predicted}}(p_i) = f(p_i) \pm \Delta f(p_i)$$

  - where p_i are the parameter(s) of interest to be determined
  - Exactly how and where this comparison happens is not relevant for now

- We *never* know the exact formula for $f^{\text{predicted}}(p_i)$, so to account for inexactness, we also quote an uncertainty $\Delta f(p_i)$

- Implies a corresponding uncertainty in extracted parameters of interest

$$\Rightarrow \qquad p_i \pm \Delta p_i$$

Outline.

- How to estimate Δf?
- How to interpret Δf, i.e., what does it actually mean?
- How to propagate Δf into Δp_i?
- What about correlations between different predictions?

How to Estimate Δf?

There are (Roughly) 3 Types of Approximations.

1. We're expanding in a (known) small quantity x and can (in principle) calculate higher-order corrections

$$f(x) = f(0) + f'(0)\,x + f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)$$

   - Example: Perturbative expansion in coupling constants

There are (Roughly) 3 Types of Approximations.

1. We're expanding in a (known) small quantity x and can (in principle) calculate higher-order corrections

$$f(x) = f(0) + f'(0)\,x + f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)$$

   - Example: Perturbative expansion in coupling constants

2. We know the limit, but don't know how to calculate corrections to it

$$f(x) = f(0) + \mathcal{O}(x)$$

   - Example: Kinematic expansion in parton showers

There are (Roughly) 3 Types of Approximations.

1. We're expanding in a (known) small quantity x and can (in principle) calculate higher-order corrections

$$f(x) = f(0) + f'(0)\,x + f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)$$

   - Example: Perturbative expansion in coupling constants

2. We know the limit, but don't know how to calculate corrections to it

$$f(x) = f(0) + \mathcal{O}(x)$$

   - Example: Kinematic expansion in parton showers

3. We don't even know a limit, and all we have is (what theorists call) a model

$$f(x) \approx \tilde{f}(x)$$

   - Example: Hadronization models

Standard Estimation Method.

Perform the expansion in slightly different ways and take the difference

- We make a variable transformation:

$$x = x(\tilde{x}) = \tilde{x} + b_0 \tilde{x}^2/2 + \mathcal{O}(\tilde{x}^3)$$

  - To lowest order $x = \tilde{x}$, so we can expand in either x or $\tilde{x}$

$$f(x) = f(0) + f'(0)\,x + f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)$$

$$f(x(\tilde{x})) = f(0) + f'(0)\,\tilde{x} + \left[f''(0) + f'(0)\,b_0\right]\frac{\tilde{x}^2}{2} + \mathcal{O}(\tilde{x}^3)$$

Standard Estimation Method.

Perform the expansion in slightly different ways and take the difference

- We make a variable transformation:

$$x = x(\tilde{x}) = \tilde{x} + b_0 \tilde{x}^2/2 + \mathcal{O}(\tilde{x}^3)$$

  - To lowest order $x = \tilde{x}$, so we can expand in either x or $\tilde{x}$

$$f(x) = f(0) + f'(0)\, x + f''(0)\, \frac{x^2}{2} + \mathcal{O}(x^3)$$

$$f(x(\tilde{x})) = f(0) + f'(0)\, \tilde{x} + \left[f''(0) + f'(0)\, b_0\right] \frac{\tilde{x}^2}{2} + \mathcal{O}(\tilde{x}^3)$$

- and conclude

$$f^{\text{predicted}} = f(0) + f'(0)\, x \pm \Delta f$$

where $\quad \Delta f = f'(0)\,(x - \tilde{x}) = f'(0)\, b_0\, \frac{x^2}{2} + \mathcal{O}(x^3)$

  - Estimated Δf is indeed $\mathcal{O}(x^2)$
  - Including the x^2 term in the prediction we get $\Delta f \sim \mathcal{O}(x^3)$

Important Caveats.

$$\Delta f = f'(0)\, b_0\, \frac{x^2}{2} + \mathcal{O}(x^3) \qquad \Delta f_{\mathrm{true}} = f''(0)\, \frac{x^2}{2} + \mathcal{O}(x^3)$$

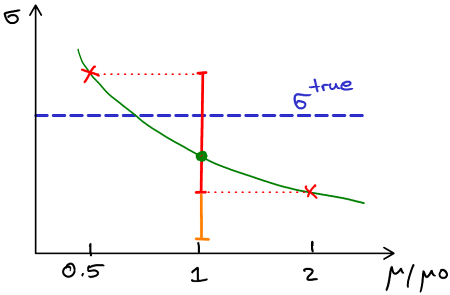
- So we effectively approximate $f''(0) \approx f'(0)\, b_0$
  - Nothing guarantees that this is a good approximation, and often it is not
  - $f''(0)$ usually has nontrivial internal structure different from $f'(0)$
  - But by default b_0 is just a constant, and the same for any f and at any order

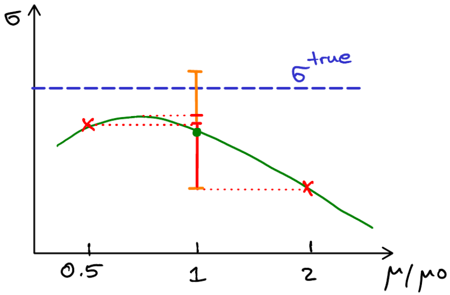
Important Caveats.

$$\Delta f = f'(0)\, b_0\, \frac{x^2}{2} + \mathcal{O}(x^3) \qquad \Delta f_{\mathrm{true}} = f''(0)\, \frac{x^2}{2} + \mathcal{O}(x^3)$$

- So we effectively approximate $f''(0) \approx f'(0)\, b_0$
  - Nothing guarantees that this is a good approximation, and often it is not
  - $f''(0)$ usually has nontrivial internal structure different from $f'(0)$
  - But by default b_0 is just a constant, and the same for any f and at any order
- Does not work if we only know the limit $f(x) = f(0) + \mathcal{O}(x)$
  - If $f(x, y)$ has more dimensions, can compare taking the limit in different ways or from different directions
- If we only have a model $f(x) \approx \tilde{f}(x)$
  - Vary model parameters or compare different models (Pythia vs. Herwig)
  - No guarantee and no way to check if this provides a good estimate

Translation to Scale Variations.





$$x \equiv \alpha_s(\mu_0)\,, \qquad \tilde{x} \equiv \alpha_s(\mu)\,, \qquad b_0 \sim \beta_0 \ln \frac{\mu}{\mu_0}$$

- Continuous choice of variable transformation
  - μ (or b_0) is *not* an actual parameter with a true value that f depends on
  - No value for it might ever capture the true result (happens regularly)
  - Uncertainty reduces at higher order because scale becomes less relevant and not because it would somehow become better known

⇒ Unfortunately so very convenient and prevalent that it is hard to overcome

Better Approach

$$f(x) = f(0) + f'(0)\,x \quad \underbrace{+ f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)}_{\text{source of the theory uncertainty}}$$

We should directly estimate $f''(0)$

- $f(x)$ is only a function of x $\Rightarrow$ $f^{(n)}(0)$ are numbers
  - Still have nontrivial internal structure (color channels, partonic channels)
- $f(x) = f(x, y)$ $\Rightarrow$ $f^{(n)}(0, y)$ are functions
  - If leading y dependence is known $\to$ expand in y and reduce to previous
- $f(x) = f(x, y_1, y_2, \ldots)$ $\Rightarrow$ $f^{(n)}(0, y_1, y_2, \ldots)$ are N-dim. functions
  - How to best estimate uncertainty due to an unknown function?

$\Rightarrow$ Will come back to this at the end

2-Point Systematics: "Herwig vs. Pythia".

Take difference of two models as the uncertainty

$$f(x) \approx \tilde{f}_1(x) \approx \tilde{f}_2(x) \qquad \Rightarrow \qquad \Delta f = \tilde{f}_2(x) - \tilde{f}_1(x) \overset{???}{\approx} \Delta f_{\mathrm{true}}$$

- Δf is small: does not mean Δf_{true} is small
  - $\tilde{f}_1(x)$ and $\tilde{f}_2(x)$ might just be equally wrong $\to$ underestimate
- Δf is large: does not mean Δf_{true} is large
  - one of $f_1(x)$ or $f_2(x)$ might just be wrong/bad $\to$ overestimate
- If both $f_1(x)$ or $f_2(x)$ can be considered equally good approximations
  - Δf may or may not give a good estimate of Δf_{true}

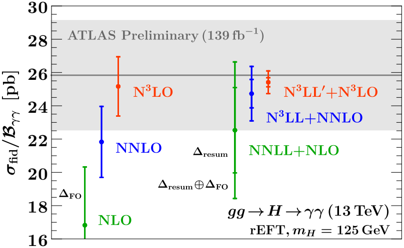
$\Rightarrow$ If this becomes a relevant source of uncertainty, best (or really only) way to proceed is to modify the analysis procedure to reduce sensitivity to it

How to Interpret Δf?

What Should Δf Actually Represent or Mean?

$$f^{\text{predicted}} = f \pm \Delta f$$

$$\Delta f \approx |f^{\text{true}} - f|$$



- We usually think of estimating possible difference to true result
  - Can only check if Δf at lower order captures next/highest known order
  - Sufficient *if* series converges well (uncertainty on uncertainty is small)
  - I tend to trust uncertainty at highest order, if lower-order uncertainties cover highest-order result (and not if they don't)
  - But: Danger of "over-tuning" lower-order results
- However, in practice almost always used as some sort of "1σ"
  - $|f^{\text{true}} - f| \leq \Delta f$ with 68% "probability"
  - But "probability" in what sense?
  - And what probability distribution?

And How Is It Distributed?

Theorist: "Do not use a Gaussian, it should be a flat distribution"

Translation: "The central value shouldn't be the most likely"

- A flat box of size Δf makes no sense (obviously too aggressive)
  - Why some theorists insist on adding theory uncertainties linearly
- How about a central flat region with some (gaussian) tails?
  - How large is the flat vs. tail region? What part does Δf cover?

And How Is It Distributed?

Theorist: "Do not use a Gaussian, it should be a flat distribution"

Translation: "The central value shouldn't be the most likely"

- A flat box of size Δf makes no sense (obviously too aggressive)
  - Why some theorists insist on adding theory uncertainties linearly
- How about a central flat region with some (gaussian) tails?
  - How large is the flat vs. tail region? What part does Δf cover?

My opinion: Use whatever distribution suits you (Gaussian, log-normal, ...)

- Until someone demonstrates that the choice actually matters
  - And if it does matter, you're so sensitive to theory uncertainties that you have much bigger problems ...
- And if a theorist complains, just do an auxiliary measurement of their true mental distribution, by asking them:

  "Which percentage of [citations on paper, monthly salary, postdoc funding, ...] are you willing to loose if the next order is outside your uncertainty? 68%? 95%?"

Correlations and How to Propagate?

Correlations.

Correlations can be crucial once several predictions are used in combination

- Prototype of extrapolation that happens in many data-driven methods

$$\underbrace{f(x)}_{\text{needed}} = \underbrace{\Big[g(x)\Big]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{f(x)}{g(x)}\right]_{\text{predicted}}}_{\text{theory uncertainties cancel}}$$

Correlations.

Correlations can be crucial once several predictions are used in combination

- Prototype of extrapolation that happens in many data-driven methods

$$\underbrace{f(x)}_{\text{needed}} = \underbrace{\Big[g(x)\Big]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{f(0) + f'(0)x \pm \Delta f}{g(0) + g'(0)x \pm \Delta g}\right]_{\text{predicted}}}_{\text{theory uncertainties cancel}}$$

  - Cancellation of theory uncertainties is often taken for granted, but obviously relies crucially on precise correlation between Δf and Δg

Correlations.

Correlations can be crucial once several predictions are used in combination

- Prototype of extrapolation that happens in many data-driven methods

$$\underbrace{f(x)}_{\text{needed}} = \underbrace{\Big[g(x)\Big]_{\text{measured}}}_{\text{measure precisely}} \times \underbrace{\left[\frac{f(0) + f'(0)x \pm \Delta f}{g(0) + g'(0)x \pm \Delta g}\right]_{\text{predicted}}}_{\text{theory uncertainties cancel}}$$

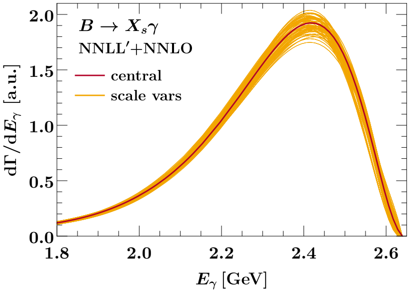
  - Cancellation of theory uncertainties is often taken for granted, but obviously relies crucially on precise correlation between Δf and Δg

- Key Issue: Correlation between Δf and Δg is not captured by our usual variation methods
  - Simultaneous (scale) variation does not imply correlation
  - Can try to come up with some theoretically motivated (but still arbitrary) correlation model
  - True correlation depends on the extent to which missing $f''(0)$ and $g''(0)$ are independent or related

Important Case: Differential Spectrum.

$$f(x, y) = f(0, y) + f'(0, y)\, x$$
$$+ \Delta f(y)$$

$$\Delta f \neq \int \mathrm{d}y\, \Delta f(y)$$



- Integral is often more precisely predicted than spectrum
  - There is a nontrivial (long-range) anticorrelation across spectrum which cancels additional (shape) uncertainty in the spectrum
- We have multiple variation estimates $\Delta f_n(y)$ which make up the band

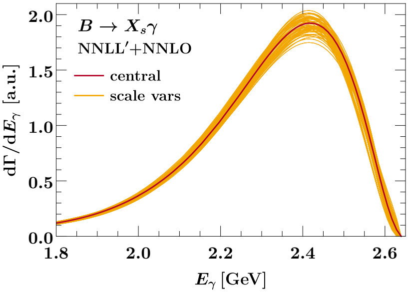
$$\Delta f(y) = \max\{|\Delta f_1(y)|, |\Delta f_2(y)|, \ldots\}$$

  - We take the envelope since they largely probe same source of inexactness
  - But envelope does not commute with integral: Taking the upper/lower edges of the band looses possible correlations and overestimates

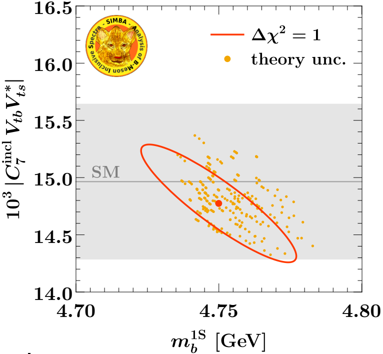
Envelope Propagation.

[Bernlochner et al., arXiv:2007.04320]

Repeat fit with varied theory inputs



$\Longrightarrow$

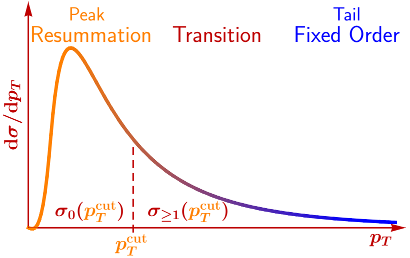


- Propagates the envelope to the final result
  - Maintains behaviour of individual variations, i.e. some form of anticorrelated shape uncertainty (which however could still be rather arbitrary)
  - But fit does not see the theory uncertainty
- How to take, interpret, and reuse envelope in fit results?
  - Should one shift the central fit value?
  - What if someone wants to use the result to predict the spectrum?

Example: Correlation Model for 2 Bins.

[Stewart, FT, arXiv:1107.2117]

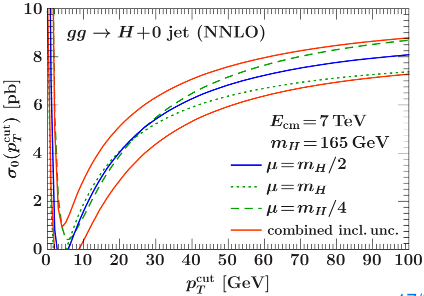
$$\sigma_{\rm tot} = \underbrace{\int_0^{p_T^{\rm cut}} dp_T \frac{d\sigma}{dp_T}}_{\sigma_0(p_T^{\rm cut})} + \underbrace{\int_{p_T^{\rm cut}}^{\infty} dp_T \frac{d\sigma}{dp_T}}_{\sigma_{\geq 1}(p_T^{\rm cut})}$$



- Scale variation fails for $\sigma_0(p_T^{\rm cut})$
- Instead, parametrize in terms of
  - yield: overall normalization
  - migration: induced by binning cut

	σ_0	$\sigma_{\geq 1}$	$\sigma_{\rm tot}$
$\theta_{\rm y}$	$\Delta_{0\rm y}$	$\Delta_{1\rm y}$	$\Delta_{0\rm y} + \Delta_{1\rm y}$
$\theta_{\rm cut}$	$\Delta_{\rm cut}$	$-\Delta_{\rm cut}$	0

  - $\Delta_{i\rm y}$ and $\Delta_{\rm cut}$ can be estimated



Theory Nuisance Parameters.

(The promise of a less ugly future)

[FT, work in progress ...]

What We Should be Doing.

Parametrize and estimate the actual source of the uncertainty: $f''(0)$

$$f(x) = f(0) + f'(0)\,x \quad \underbrace{+ f''(0)\,\frac{x^2}{2} + \mathcal{O}(x^3)}_{\text{source of the theory uncertainty}}$$

- We typically know a lot about the general structure of $f''(0)$ even without explicitly calculating it
  - Color structure, partonic channels, kinematic structure, ...
  - All we want is an uncertainty estimate, so it is sufficient to consider dominant contributions or limits
- Parametrize $f''(0)$ and treat the remaining unknown parameters as "theory nuisance parameters" (TNPs)
  - Figure out allowed range based on theory arguments
  - Best case: Parameters are numbers
  - More generally, one or more unknown functions

Advantages of Theory Nuisance Parameters.

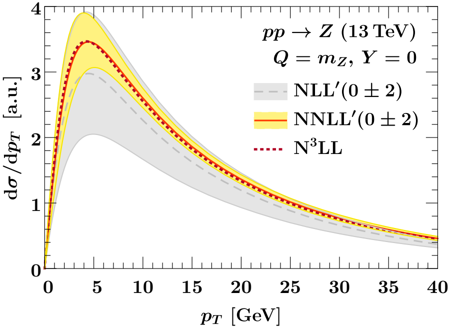
TNPs are genuine parameters with a true but unknown or uncertain value

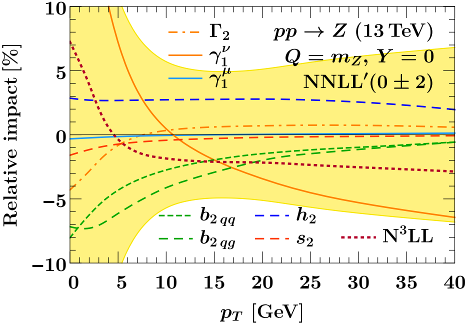
- Renders the whole problem much more well-defined
- We get all benefits of truly parametric uncertainties
  - ✓ Encode correct correlations, straightforward to propagate everywhere
  - ✓ Can be constrained by measurements (auxiliary and/or primary)
- There will typically be several parameters
  - Much safer against accidental underestimate of any one parameter
  - Total theory uncertainty becomes Gaussian due to central-limit theorem
- Can even lead to reduced theory uncertainties
  - Can fully exploit partially known higher-order information
  - Can also reduce theory uncertainties at a later time

Price to pay

- Predictions become quite a bit more complex
  - Need to implement complete next order in terms of unknown parameters

Example: Z p_T Spectrum.

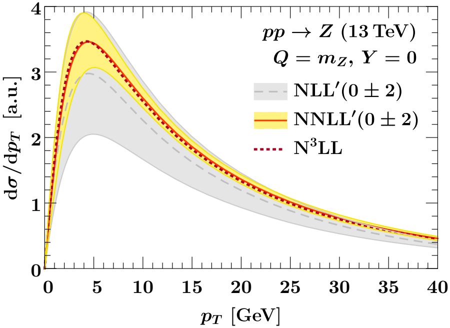


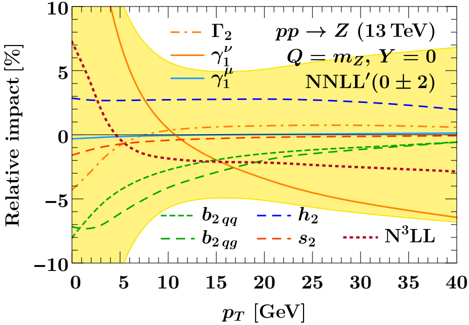


$$f(x, p_T) = \exp\Big\{\sum_i \Big[f_i(x)\Big]\, g_i(p_T)\Big\} + \mathcal{O}\Big(\frac{p_T^2}{m_Z^2}\Big)$$

- Here, leading p_T dependence factorizes, $g_i(p_T)$ are known exactly

Example: $Z\ p_T$ Spectrum.

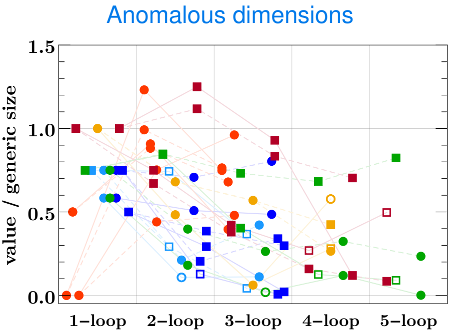


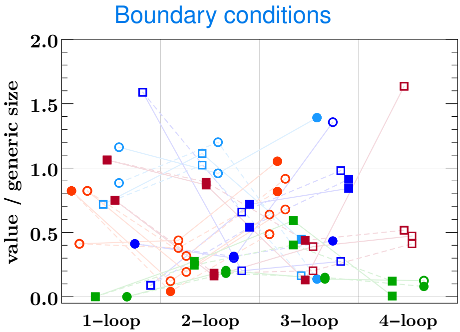


$$f(x, p_T) = \exp\left\{\sum_i \left[f_i(0) + f_i'(0)\,x + f_i''(0)\frac{x^2}{2}\right] g_i(p_T)\right\} + \mathcal{O}\left(\frac{p_T^2}{m_Z^2}\right)$$

- Here, leading p_T dependence factorizes, $g_i(p_T)$ are known exactly
- Problem reduces to parametrizing $f_i''(0)$ which are numbers
  - Correlations in p_T spectrum are fully captured ✓
  - Illustration: Show $\theta_i = (0 \pm 2)\theta_i^{\text{true}}$ with known θ_i^{true} at this order

Estimating Size of TNPs.





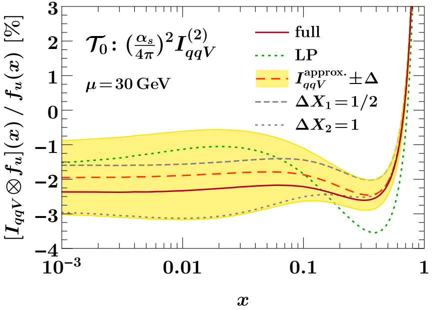
- Possible to estimate the typical size of TNPs (when they are numbers)
  - Can construct a general estimator based on known structure of perturbation theory (basically leading color and n_f dependence)
  - Shown are coefficients of many known perturbative series divided by corresponding estimate at each order

Example of a Functional TNP.

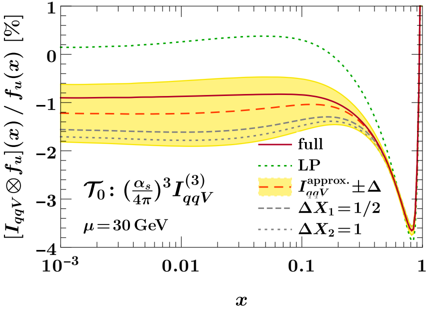
Remaining challenge is when TNPs are genuine functions

- **Strategy:** Parametrize by exploiting known functional dependence and/or expanding in known limits
- **Example:** Beam function matching coefficients depend on parton momentum fraction x (similar to splitting functions)
  - Can construct a parametrization based on expanding around $x \to 1$
    [Billis, Ebert, Michel, FT, arXiv:1909.00811]

NNLO (full was known)



N^{3}LO (full was not yet known)

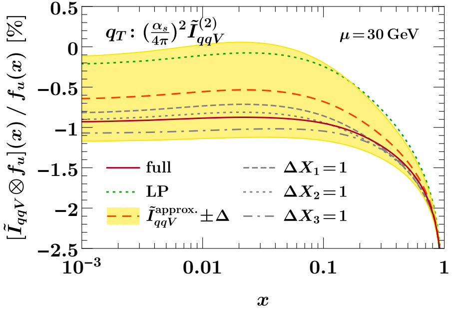


Example of a Functional TNP.

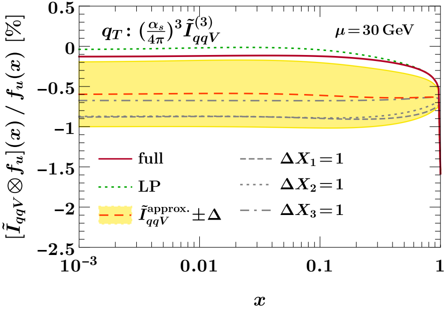
Remaining challenge is when TNPs are genuine functions

- **Strategy:** Parametrize by exploiting known functional dependence and/or expanding in known limits
- **Example:** Beam function matching coefficients depend on parton momentum fraction x (similar to splitting functions)
  - Can construct a parametrization based on expanding around $x \to 1$
    [Billis, Ebert, Michel, FT, arXiv:1909.00811]

NNLO (full was known)



N^{3}LO (full was not yet known)



Summary.

Theory uncertainties are indeed ugly business

- Be aware of limitations of current methods like scale variations
  - Not particularly reliable
  - Most severe limitation is the lack of proper correlations

- Some might say that the best way is to avoid theory uncertainties
  - But “avoiding” often secretly means “canceling” them, which relies on correlations, so we’re right back to where we started

We can make progress when we have an actual expansion

- Parametrize the known unknown: theory nuisance parameters
  - A paradigm change, but the obvious way forward (at least to me)
  - Any feedback is most welcome ...

Acknowledgments.

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 101002090 COLORFREE)

erc

European Research Council

Established by the European Commission