

# **Unfolding Methods**

## in Particle Physics for Quantum Information Observables

### Michele Pinamonti

(INFN Trieste, University of Udine, ICTP Trieste)

Quantum Observables for Collider Physics GGI Florence, 10 Nov 2023

## What is unfolding about?

- Unfolding is:
  - removal of "imperfect measurement device" effects from observed distribution, to extract (our best-guess of) underlying true distribution



- In the case of **collider physics** *"imperfect measurement device" effects* include:
  - detector resolution
  - acceptance / efficiency corrections
  - parton-to-particle level evolution
  - final-state reconstruction or combinatorial background

Interesting readings: <u>slides</u> by Glen Cowan, <u>proceedings</u> by Stefan Schmitt (unbiased selection)



## Why does unfolding matter here?

#### Testing Bell inequalities at the LHC with top-quark pairs

M. Fabbrichesi<sup>†</sup>, R. Floreanini<sup>†</sup>, and G. Panizzo<sup>\*</sup>

https://arxiv.org/abs/2102.11883



We find that, under such conditions, the null hypothesis and the violation of Eq. (3) can be assessed at the 98% CL with present Run II luminosity. Moreover, after rescaling this result by the projected luminosity of the LHC full Run III, we expect that it will be possible to test the violation at the 99.99% CL ( $4\sigma$  significance).

CHSH on fixed	axes, $\sqrt{2}  - C_{rr} + C_{nr}$
High- $p_T$	Significance for $> 2$
Selection	$[3  ab^{-1}]$
Weak	$1.7 \sigma$
Intermediate	$1.8 \sigma$
Strong	$1.0\sigma$

### Quantum tops at the LHC: from entanglement to Bell inequalities

Claudio Severi<sup>1</sup> <sup>a</sup>, Cristian Degli Esposti Boschi<sup>3,4</sup> <sup>b</sup>, Fabio Maltoni<sup>2,4,5</sup> <sup>c</sup>, and Maximiliano Sioli<sup>2,4</sup> <sup>d</sup>

https://arxiv.org/abs/2110.10112

• Main difference <sup>(\*)</sup> tracked down to come from different evaluation of resolution effect from tt reconstruction + unfolding to parton level

(\*): other details matter as well, like definition of observable, exact selection applied, simultaneous fit to several bins...

## To unfold or not to unfold?

- As we will see, unfolding can bring technical and/or conceptual problems
- <u>Alternative approaches</u>:
  - Detector-level template fit:
    - measurement extracted from comparison of templates for each tested hypothesis
    - templates obtained either via MC simulation or through "folding"

m(ff) < 380 GeV</p>

Reconstructed cose

40000

30000

20000

10000

- Calibration curve:
  - measurement extracted at detector level
  - obtained value corrected via calibration curve



Reconstructed D

-0.6

340 < m(tt) < 380



Invaria

## Advantages of "proper" unfolding

- So, why doing unfolding?
  - the key is **reusability**!



#### • Some of the other possible advantages:

- **combination** of more information (more channels, signal regions, bins...)
- improve precision by inserting well-motivated **bias** (regularization)

0 ...

## How does unfolding work: the binned case



- The unfolding problem can be essentially reduced to a **response-matrix**-inversion problem
- Can be done to extract:
  - total-phase-space or fiducial-phase-space cross-sections
  - cross-sections vs. variable defined at particle-level or at parton-level

## Naive unfolding: matrix inversion or maximum likelihood



• **Matrix inversion** gives a simple solution to the unfolding problem:

$$t_j = \sum_i M_{ij}^{-1} d_i$$

• Numerically, the same result can be obtained via **maximum-likelihood** principle:

$$L(t_{j}) \sim \exp[-\sum_{i} (d_{i} - \sum_{ij} M_{ij} t_{j})^{2}]$$

- $\circ$   $T_i$  here treated as free-floating parameters
- best-fit values of  $T_i$  gives the unfolded distribution
- **Note**: maximum-likelihood can be used for unfolding with non-square M matrix ( $N_{reco-bins} > N_{truth-bins}$ )

## What about background and systematics?

- Background:
  - background needs to be subtracted from observed data *before* unfolding

$$t_j = \sum_i M_{ij}^{-1} (d_i - b_i)$$

- Systematic uncertainties:
  - where are they entering the game?
    - response matrix M
    - background estimation b



## Naive unfolding issues...

- Matrix-inversion (or maximum-likelihood) can produce results with large "oscillations"
  - such **amplification of statistical fluctuations** is is related to off-diagonal elements in M



## Ways to overcome oscillations

- Possible ways to mitigate to statistical fluctuation amplification in unfolding:
  - "let's ignore the off-diagonal elements in M"  $\rightarrow$  bin-by-bin correction factors
  - *"let's reduce off-diagonal elements in M"* → experimental resolution improvement
     → binning optimisation
  - "let's add more information"

 $\rightarrow$  maximum-likelihood unfolding with  $N_{reco-bins} > N_{truth-bins}$ 

- "let's just dump these oscillations"
  - $\rightarrow regularisation$



## The regularization idea

- **Regularization** in unfolding means adding **external constraints** in order to dump oscillations
- Simple example the maximum likelihood formalism:

$$L(t_{j}) \sim \exp[-\chi^{2}]$$
where
$$\chi^{2} = \sum_{i} (d_{i} - \sum_{ij} M_{ij} t_{j})^{2} / d_{i} + \tau^{2} \sum_{j} (t_{j} - t_{j}^{0})^{2}$$

• More in general (Tikhonov regularization):

$$\chi^{2} = \chi_{0}^{2} + \tau^{2} \chi_{R}^{2} \quad \text{with} \quad \chi_{R}^{2} = (t - t^{0})^{T} R^{T} R (t - t^{0})$$

Tikonov factor controlling regularization strength

regularization condition matrix (more often used **2<sup>nd</sup> derivative constraint**)



## **Regularization in action**







## **Unfolding checks: variance vs. bias**

- **Regularisation**, in any form, introduces a **bias**:
  - *external information* is added to the system, in the form of prior probability distribution of parameters or relationships between them (*e.g. constraint on distribution smoothness*)
- Bias is usually quantified with a number of **stress-tests**:
  - basic (but pretty general) example:
    - pseudo-data produced by folding (with M) alternative truth-distribution t<sub>0</sub>', to obtain a detector-distribution d'
    - this d' unfolded (keeping M and regularization built with nominal model) to obtain t'
    - *difference* t'-t<sub>0</sub>' quantifies bias
- Strength of regularization usually determined by minimizing bias and at the same time minimizing variance (i.e. statistical uncertainty, i.e. expected statistical fluctuations)

## **Iterative Bayesian Unfolding (IBU)**

• Uses Bayes theorem iteratively:

**Systematics**:



- prior based on theoretical prediction in first iteration
- following iterations use result of previous ones as prior



### $p_1(T|D) \propto \mathcal{L} \cdot \pi(T)$ $p_2(T|D) \propto \mathcal{L} \cdot p_1(T|D)$ $p_3(T|D) \propto \mathcal{L} \cdot p_2(T|D)$

## Interlude: Profile Likelihood Fit (PLF)

- An approach to *properly* include systematic uncertainties in maximum-likelihood fits:
  - include systematic uncertainties as unknown parameters in the model:
    - nuisance parameters modifying expectations in a parametric way
  - prior probabilities on values of nuisance parameters to reflect limited knowledge



#### 15

## **Nuisance parameters and systematic uncertainties**

normal distribution



## **Profile Likelihood Unfolding (PLU)**

- Unfolding can be handled with the profile-likelihood method
- How does it work?
  - profile-likelihood unfolding
  - basic idea very simple:
    - reco distribution is a function of the differential cross-section, i.e.
       the *N* parameters of interest (POIs), where *N* is the number of truth-distribution bins:

$$L(n \mid \theta, k, \sigma_{l}, ..., \sigma_{N}) = \prod_{i} P(n_{i} \mid S_{i}(\theta, k, \sigma_{l}, ..., \sigma_{N}) + B_{i}(\theta, k)) \times \prod_{j} G(\theta_{j})$$
  
or

 $L(n \mid \boldsymbol{\theta}, \boldsymbol{k}, \boldsymbol{\mu}_{l}, \dots, \boldsymbol{\mu}_{N}) = \prod_{i} P(n_{i} \mid S_{i}(\boldsymbol{\theta}, \boldsymbol{k}, \boldsymbol{\mu}_{l}, \dots, \boldsymbol{\mu}_{N}) + B_{i}(\boldsymbol{\theta}, \boldsymbol{k})) \times \prod_{j} G(\boldsymbol{\theta}_{j})$ 

with  $\mu$  = **signal strengths** for the different truth bins ( $\sigma_k = \mu_k \sigma_k^{th}$ )

- **unfolding** = **maximum-likelihood fit** to find best-fit values of  $\sigma_{1},...,\sigma_{N}$
- systematics included in the same (*natural*) way as usual PLF

## Profile likelihood unfolding in action

• CMS Top mass from tt+1jet differential cross-section

JHEP 07 (2023) 077



## **Profile likelihood unfolding in action (II)**

• ATLAS ttW differential cross-section measurement

ATLAS-CONF-2023-019



\*: "migration matrix" actually (without acceptance and efficiency correction parts)

## **Unfolding for Entanglement** & Bell inequalities - The ideal case

•  $tt \rightarrow dilep case$ 





## **Reality - so far...**

• Why didn't do unfoldind for the ATLAS result?



## Where could profile-likelihood unfolding help?

- When **background** is large:
  - main use-case:  $H \rightarrow WW$  !
  - can have in-situ background constraints
- To constrain systematic uncertainties:
  - in case we get dominated by systematics that we *are allowed* to constraint
    - need  $N_{reco-bins} > N_{truth-bins}$ (which, by the way, mitigates oscillations...)
- To sensibly **combine** different channels:
  - different decay channels
  - event categories with different purities or resolution...
  - combination giving single unfolded distribution to be used to extract the result(s)



• ..

0

## To regularize or not to regularize?

- Results can look "unphysical", even if they are "correct":
  - estimated  $t_i$  are strongly (anti-)correlated
  - zero bias by definition:
    - by folding back the obtained distribution, we re-obtain he reco-data
    - by unfolding an alternative reco-pseudo-data, we obtain the correspondent alternative truth-distribution
- Then, **do we really need to use regularization**, or can we just use unregularized unfolding?
  - extraction of quantities from unfolded distribution can/should properly take into account correlations!
  - however: important to consider numerical precision of result when strong (anti-)correlations are involved...



## Is unfolding model-independent?

- Is unfolding safer than calibration curve?
- Is unfolding procedure independent on underlying physics we want to test?
  - at least the unregularized one?
- Actually, **response matrix** built with SM Monte Carlo...
  - different hypotheses for underlying physics could in general change it!
  - relevant specific case:
    - would a hypothetical "entanglement-off" MC sample lead to the same response matrix?

$$t_j = \sum_i M_{ij}^{-1} (d_i - b_i)$$

• In addition: background could also depend on underlying physics...

## **Other kinds of unfolding**

- Other *classical* unfolding methods exist:
  - Fully Bayesian Unfolding
  - SVD unfolding
  - 0 ...
- In addition, **new unfolding methods** being developed:
  - **OmniFold** <u>Phys. Rev. Lett. 124, 182001 (2020)</u> (Nachman, Andreassen, Komiske, Metodiev, Thaler)
    - unbinned unfolding
    - based on iterative machine-learning reweighting
  - Unfolding via Quantum Annealing
     J. High Energ. Phys. 2019, 128 (2019) (Cormier, Di Sipio, Wittek)
    - likelihood-based regularized unfolding on a quantum computer



Quantum Processing Unit

Unfolded Distribution

### **Time for discussion!**