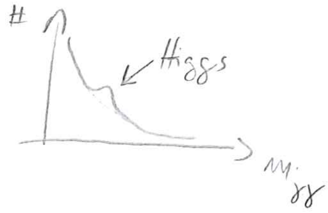
Unfolding

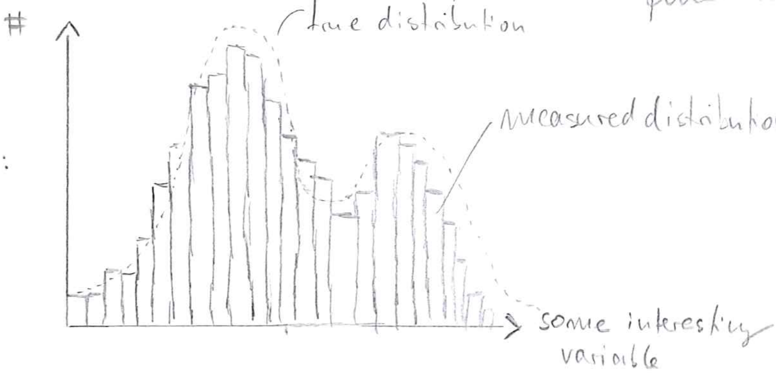
Particle physics:
1. search for new physics
2. get better understanding of known physics

For accomplishing this (experimentalists):
get estimation of probability distributions of ...
- energy/momentum
- scattering angles
- particle mass / invariant mass of objects
- ...



using $\underbrace{\text{large data sets}}_{\text{statistical fluctuations}}$ recorded by $\underbrace{\text{very complex machine}}_{\text{systematic fluctuations}}$.

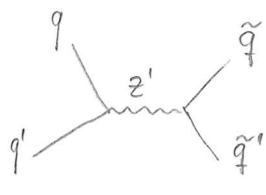
This leads to:



↕ stat. fluctuations
↔ bin migration (finite resolution)
↓ limited acceptance/efficiency
↙ non-linear response of detector

Searching for new physics

- Let's say a specific theory is to be tested, e.g.

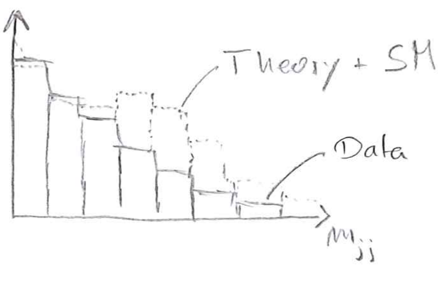


with $m_{Z'}$ fix, e.g. $m_{Z'} = 1\,\text{TeV}$

(similar for DM)

- You have: theory & data of measurement
  - theory ↓ "truth level"
  - data of measurement ↓ "reconstruction level"
- Does theory agree with measured data?

Theory Truth	Showering & Detector effects ⟹	Theory Reco
MC Simulation	„an e^- with these parameters will look like this in ATLAS...", $q \to$ jet, ...	Comparable to recorded Data



Pro's & Con's of this "folding"

Pro's

+ super quick
+ super easy
+ super convenient for experimentalists
+ analyses can gauge their sensitivities using pre-defined reference models

Con's

- doesn't allow for comparisons with other experiments
- Can be done only for limited amount of theories
- super inconvenient for theorists

The solution: Unfolding

1. Start with measured distributions (Reco-level)
2. Remove detector effects (↔, ↓, ↙ from earlier)
3. Compare unfolded data to theories/other experiments on "truth"-level

→ deals with all the Con's from before, but is somewhat complicated...

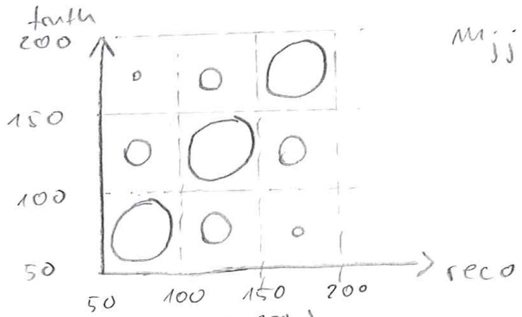
Naïve approach:

$$\boxed{\text{Detector level distribution}} = \text{Detector Effects} \times \boxed{\text{Truth level distri.}}$$

n-Tuple/Vector = Matrix × n-Tuple/Vector

"totally a no-brainer, just inverse the matrix!"

Could work like this:



get this from MC simulation

Now a measured vector $N(R) = \begin{pmatrix} n_{150}^{200} \\ n_{100}^{150} \\ n_{50}^{100} \end{pmatrix}$ can be unfolded into a truth level vector $N(T)$

<u>However</u>: isn't this more of a probabilistic problem?

$$N(R)_i = P(R_i | T_j) \times N(T_j)$$

↳ unfolding then requires $P(T_j | R_i)$... ?

-> <u>Bayes Theorem</u>:

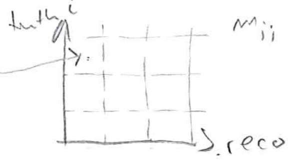
$$P(T_i | R_j) = \frac{P(R_j | T_i) \cdot \overbrace{P_0(T_i)}^{\text{Prior for } T_i}}{\underbrace{\sum_{k=1}^{N_{truth}} P(R_j | T_k) \cdot P_0(T_k)}_{\text{prior for } R_j}}$$

=> Bayesian Unfolding:

$$n(T_i) = \sum_j P(T_i | R_j) \cdot n(R_j)$$

↳ measured data

↳ requires $P(R_j | T_i) = \dfrac{N_{ij}}{N(T_i)}$



↳ requires $P_0(T_i) = \dfrac{N(T_i)}{N(T_{tot} + \underbrace{N_{Fakes}}_{})}$

↳ truth didn't survive cut, but reco did

-> take new bin into consideration:

$$P(T_i | R_j) = \frac{P(R_j | T_i) \cdot P_0(T_i)}{\sum_{k=1}^{N_{truth}+1} P(R_j | T_k) P_0(T_k)}$$

Opposite effect: reco event didn't survive, but truth did -> inefficiency!

↳ $n(T_i) = \dfrac{1}{\varepsilon_i} \sum_j P(T_i | R_j) n(R_j)$

Uncertainties: "It's difficult..."

- MC stats finite: unfolding matrix has uncertainties
  -> vary matrix 100-1000 times within uncertainties and repeat unfolding
    ↳ spread is uncertainty. Works also with sys. uncert.
- Data also has statistical uncertainty
  -> not dominating except for using <u>iterative</u> unfolding
    ↳ use $n(T_i)$ as prior for following unfolding iteration
      ↳ error propagation (each time)