School of Statistics 2024, Carry-le-Rouet

David Rousseau, IJCLab

ML for Higgs physics tutorial

Using ML to see the Higgs Boson
Using Boosted Decision Tree
Introduction to tutorial

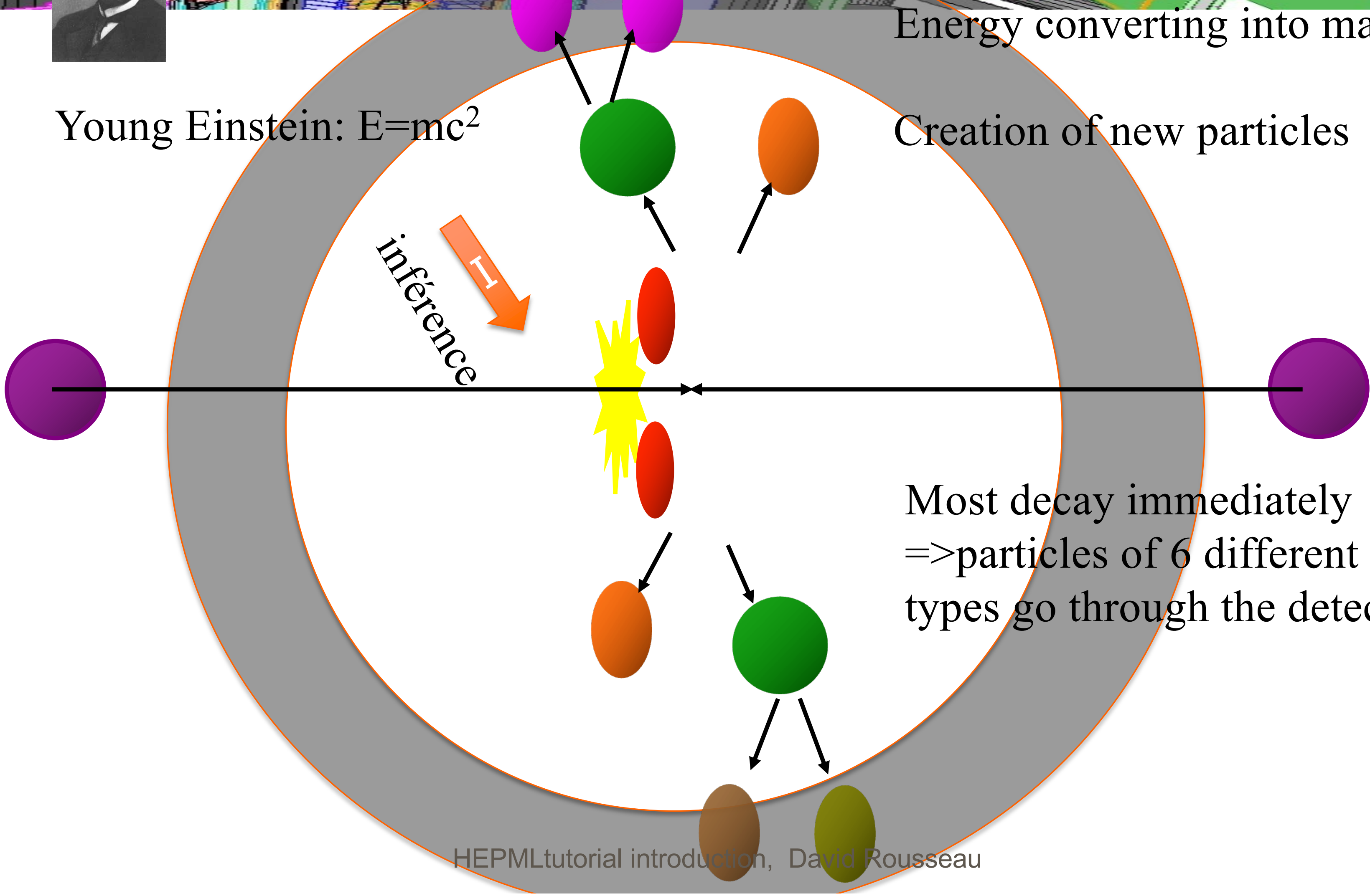
Seeing the Higgs boson

Proton collisions

Young Einstein: $E=mc^2$

Energy converting into mass

Creation of new particles



Most decay immediately
=>particles of 6 different
types go through the detector

Two fundamental entities

- « Events » :
  - All measurements from one proton collision
  - List of particles with their properties
  - Derived quantities
  - ➔ML to help select interesting events « Signal » with respect to « Background »
- « Particles »:
  - Extracted from an event
  - Jet, lepton, photon Missing ET
  - ➔ML to help identifying particles, regressing properties

Before observation, all was known about the Higgs boson, except its mass

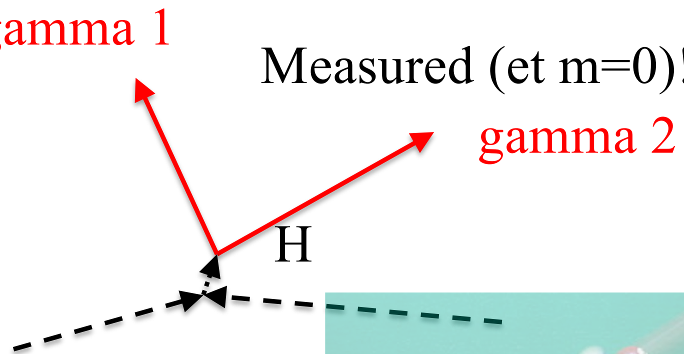
Probabilités de désintégration prédites pour une masse de 125 GeV

$H \rightarrow bb$	58%
$H \rightarrow WW^*$	21%
$H \rightarrow \tau^+\tau^-$	6.4%
$H \rightarrow ZZ^*$	2.7%
$H \rightarrow \gamma\gamma$	0.2%

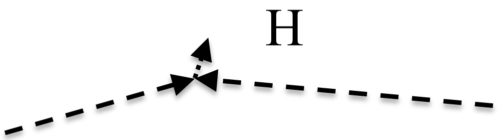
$E=mc^2$

Einstein en 1905

$E^2=p^2+m^2$



Measured (et m=0)!



H, before decay

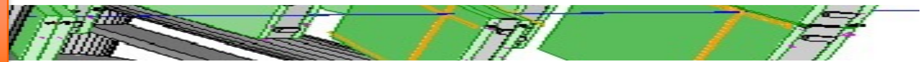
$m_H{}^2=E_H{}^2-p_H{}^2$

H,Just after decay

$E_H=E_{g1}+E_{g2}$

$\vec{p}_H=\vec{p}_{g1}+\vec{p}_{g2}$

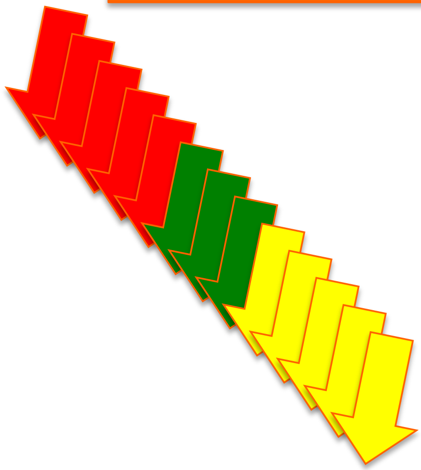
⇨we get m_H!



10^{14} collisions / year

Trigger: fast rough selection

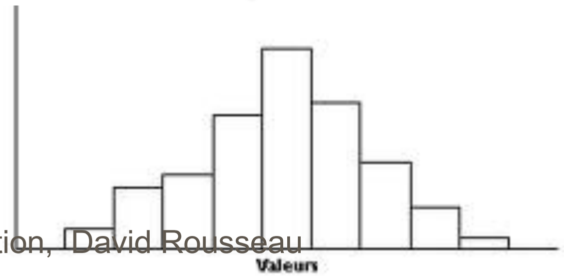
10^{9} events on disk

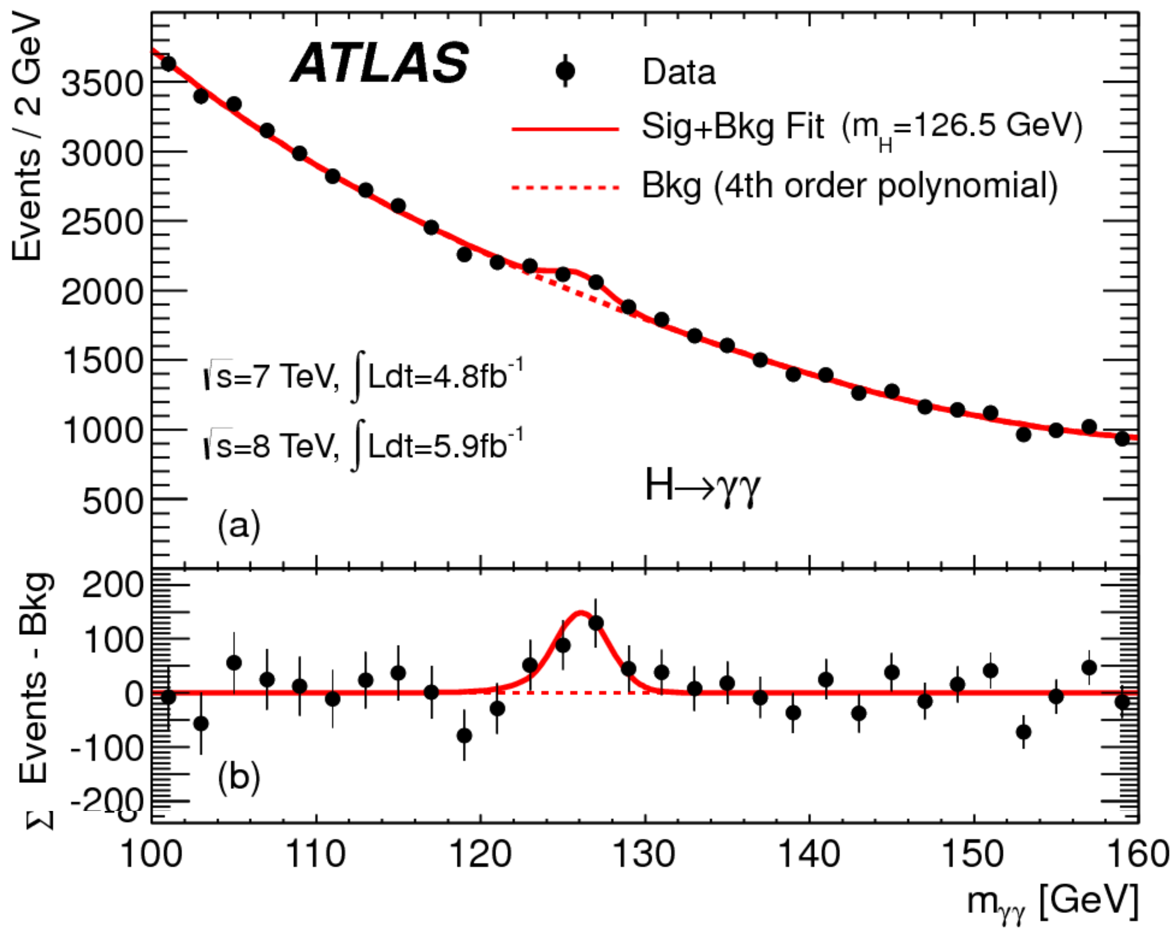


Offline selection

10^{5} events with 2 photons

Mass calculation
→histogram

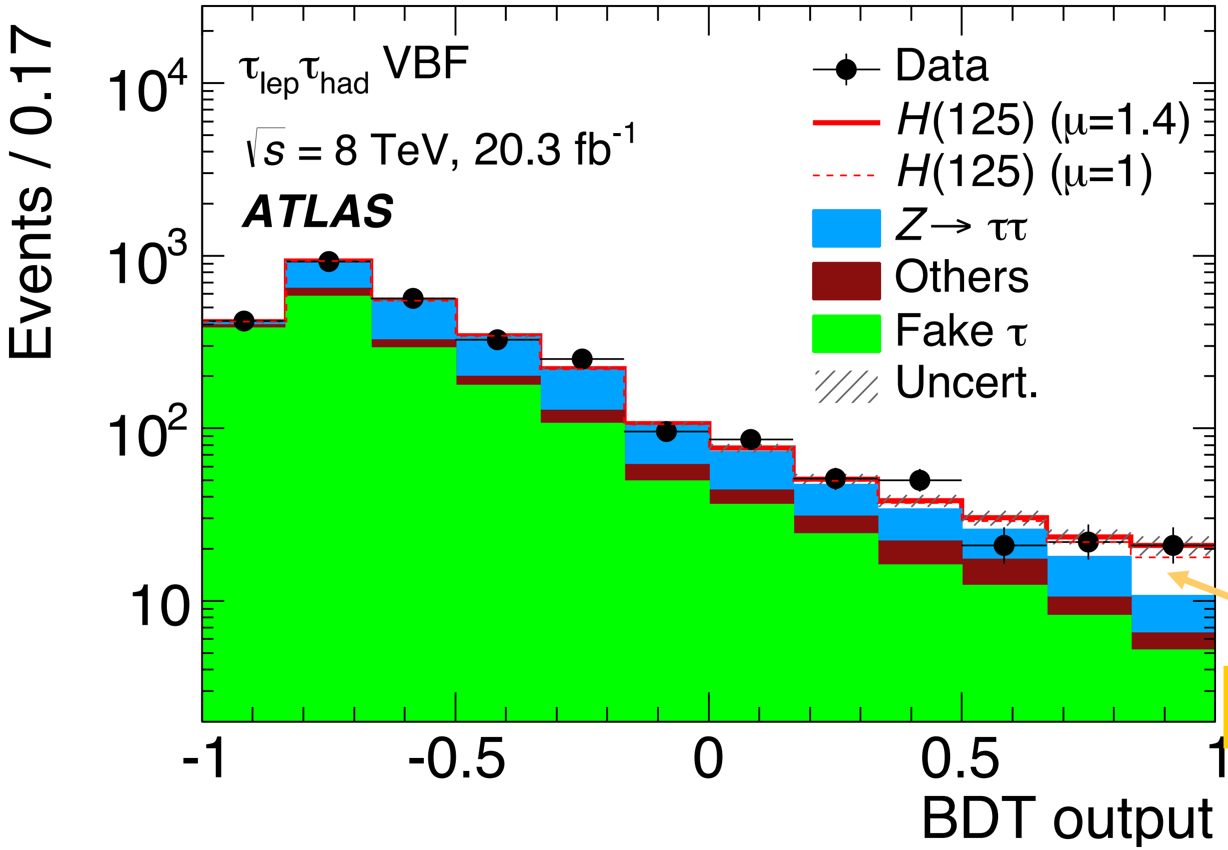


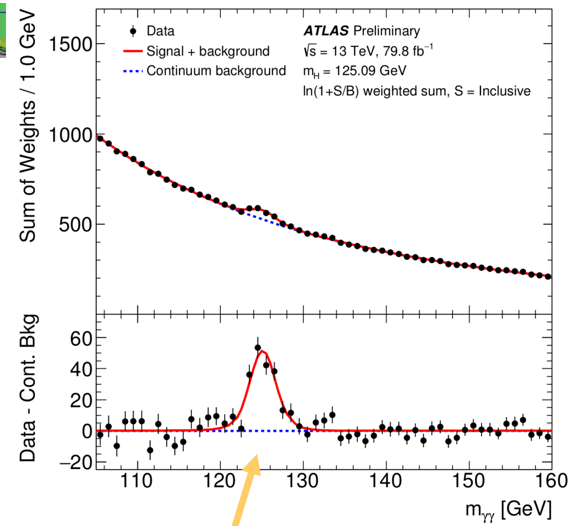

ATLAS
Data
Sig+Bkg Fit (m_H=126.5 GeV)
Bkg (4th order polynomial)
√s=7 TeV, ∫Ldt=4.8fb-1
√s=8 TeV, ∫Ldt=5.9fb-1
H→γγ
(a)
(b)
Events / 2 GeV
Σ Events - Bkg
m_γγ [GeV]

Classifier in Higgs Physics

JHEP 04, 117 (2015) 1501.04943

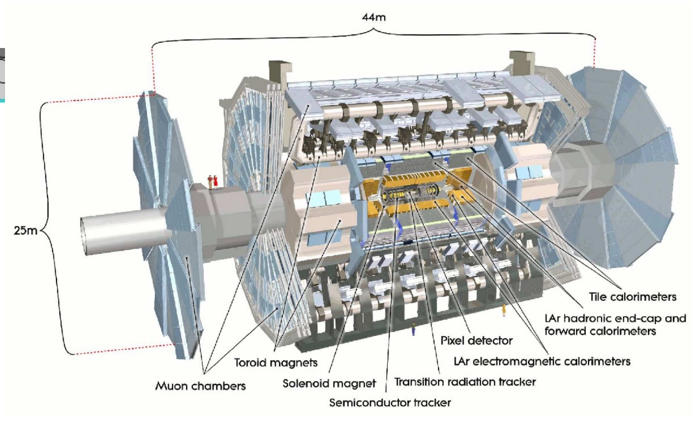
BDT using ~dozen of high level variables

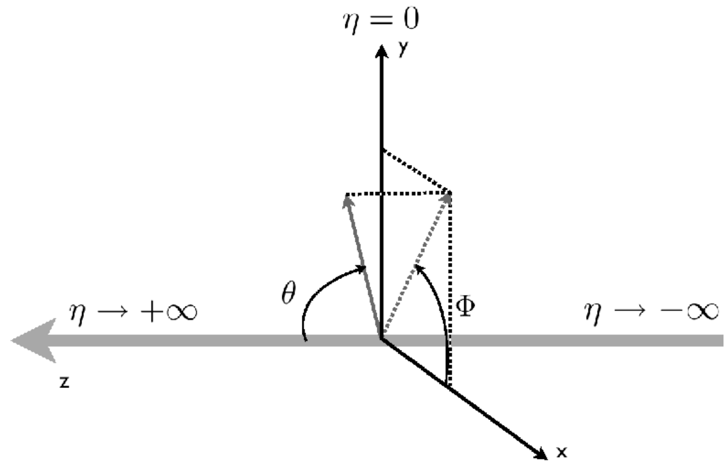




Higgs evidence

Coordinates

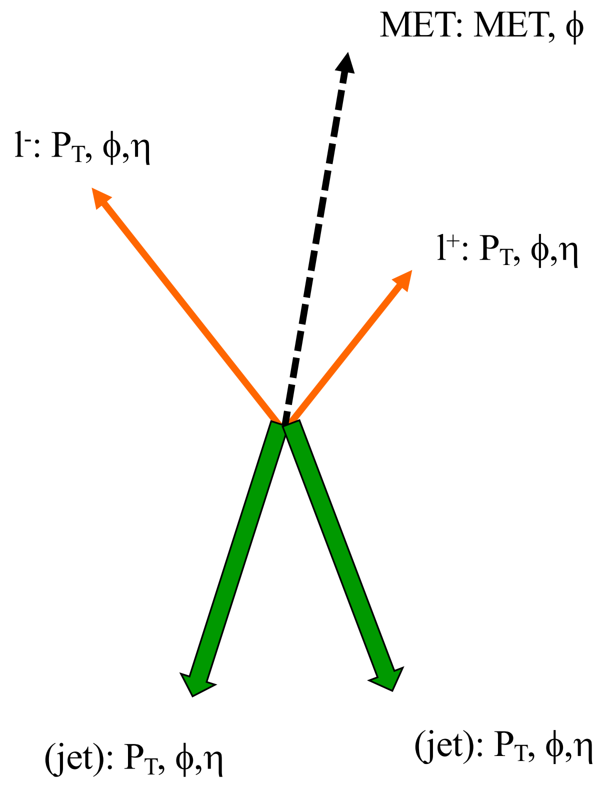




- P : momentum
- E : energy =sqrt(P^2+M^2)~P because P>>M
- Angles (cylindrical)
  - ϕ : azimuth angle]-π, +π]
  - θ : dip angle [0, +π]
  - η : eta, pseudo-rapidity = -log(tan(θ/2)), ~[-5,5]
- P_T : =P sin(θ) : transverse momentum
- ME_T : Missing Transverse Energy = $-\Sigma_{\text{all particles}} \vec{P_T}$: estimator of transverse momentum of neutrinos

Tutorial dataset H→WW

- One of the Higgs Discovery channel
- $H \to W^+(\to l^+ \nu)\ W^+(\to l^- \nu)$
  - ➔2 leptons of opposite charge
  - Neutrinos undetected ! => Missing Transverse Energy
  - No invariant mass peak!
- Background :
  - Other processes leading to $W^+(\to l^+ \nu)\ W^+(\to l^- \nu)$



Event weighting

Absolute normalisation

- Say you are doing an experiment at the LHC
- You are looking for a particular type of event
- How many do you expect ?
- $N^{prod} = L * \sigma(\theta)$
  - N^{prod}= number of produced events (before detector effect)
  - L « integrated luminosity » : for example 138 fb^{-1} for LHC data taking at 13TeV center of mass energy in 2015-2018 prop number of proton collisions
    - 1 barn is 10^{-28} m^2
    - proportional to the total number of proton collision
  - $\sigma(\theta)$: cross-section (in barn), can be calculated from first principles and θ parameters from nature (electric charge, higgs boson mass etc…)
- $N^{exp} = L * \sigma(\theta) * \varepsilon$
  - N^{exp}= number of expected events (actually counted in the detector). N^{exp} is a real number. The actual number of observed event will follow Poisson (N^{exp})
  - ε : efficiency, probability to detect a produced event (1. if perfect detector).
    - Measured on simulation (calibrated on data)
    - Can be product of many terms like: ε trigger $*$ ε acceptance $*$ ε lepton $*$ ….

Simple Event Counting Experiment

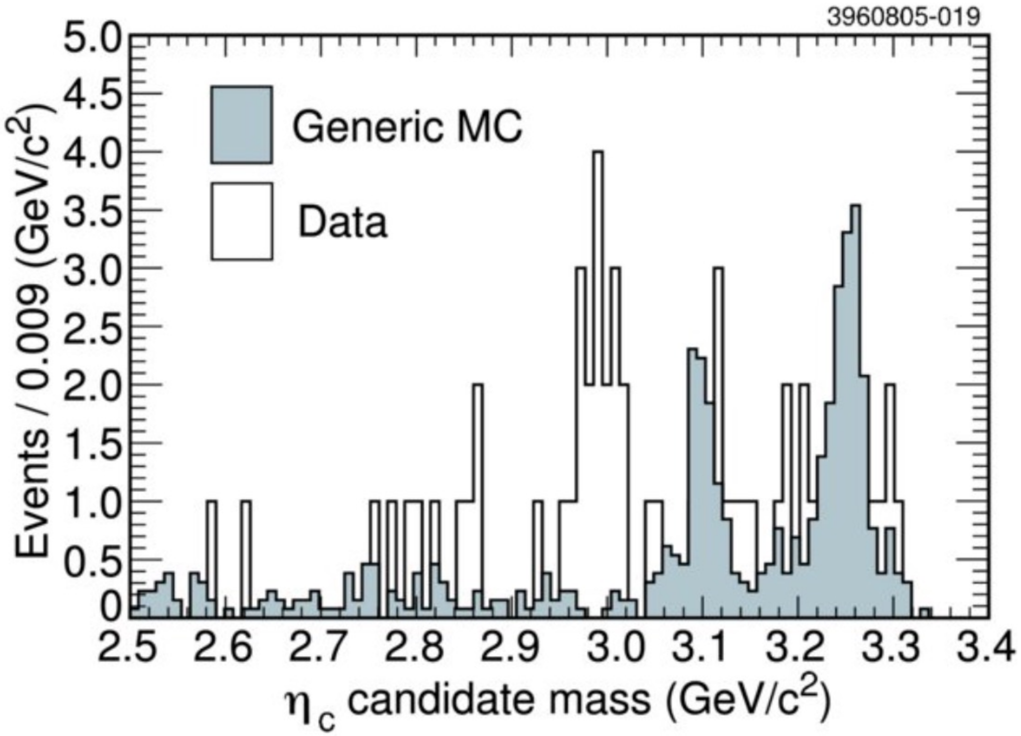
- One signal, we have some estimate of $\sigma_{sig}(\theta)$ but we actually want to assess its existence (exp==expected)
  - N^{exp}_{sig}==s=L * $\sigma_{sig}*\varepsilon_{sig}$
- one well-known background :
  - N^{exp}_{bkg}==b=L * $\sigma_{bkg}*\varepsilon_{bkg}$
- N^{exp}=s+b
- We do the experiment and count N^{obs} events
- Hence we measure:
  - $\sigma_{sig}=(N^{obs}- b)/(L*\varepsilon_{sig})$
  - $\sigma_{sig}=(N^{obs}- L*\sigma_{bkg}*\varepsilon_{bkg})/(L*\varepsilon_{sig})$
- Key inputs : ε_{sig} ε_{bkg} determined from simulated datasets

Weights for overall normalisation

- $b = L * \sigma_{bkg} * \varepsilon_{bkg}$
- We measure on simulation : $\varepsilon_{bkg} = N_{bkg\ pass}/N_{bkg\ total}$
  - with $N_{bkg\ pass}$, number of events passing some criteria e.g. momentum of the two photons greater than 25 GeV, BDT score above 0.8 etc…
  - So $b = L * \sigma_{bkg} * N_{bkg\ pass}/N_{bkg\ total}$
- We can define an event weight : $w_i = L * \sigma_{bkg} / N_{bkg\ total}$
- And then simply: $b = \Sigma_{pass}\ w_i$
- Beware : if I take an unbiased subset of x% of dataset, I need to scale the weights by 1/x, so that
- $b^{subset} = \Sigma^{subset}{}_{pass}\ w^{subset}{}_i = (1/x) * \Sigma^{subset}{}_{pass}\ w_i \sim b$

Data / MC histo comparison

- Then one can histogram directly any quantity (using the weights) and it is normalised correctly to the real data
- By convention, real data is almost never weighted

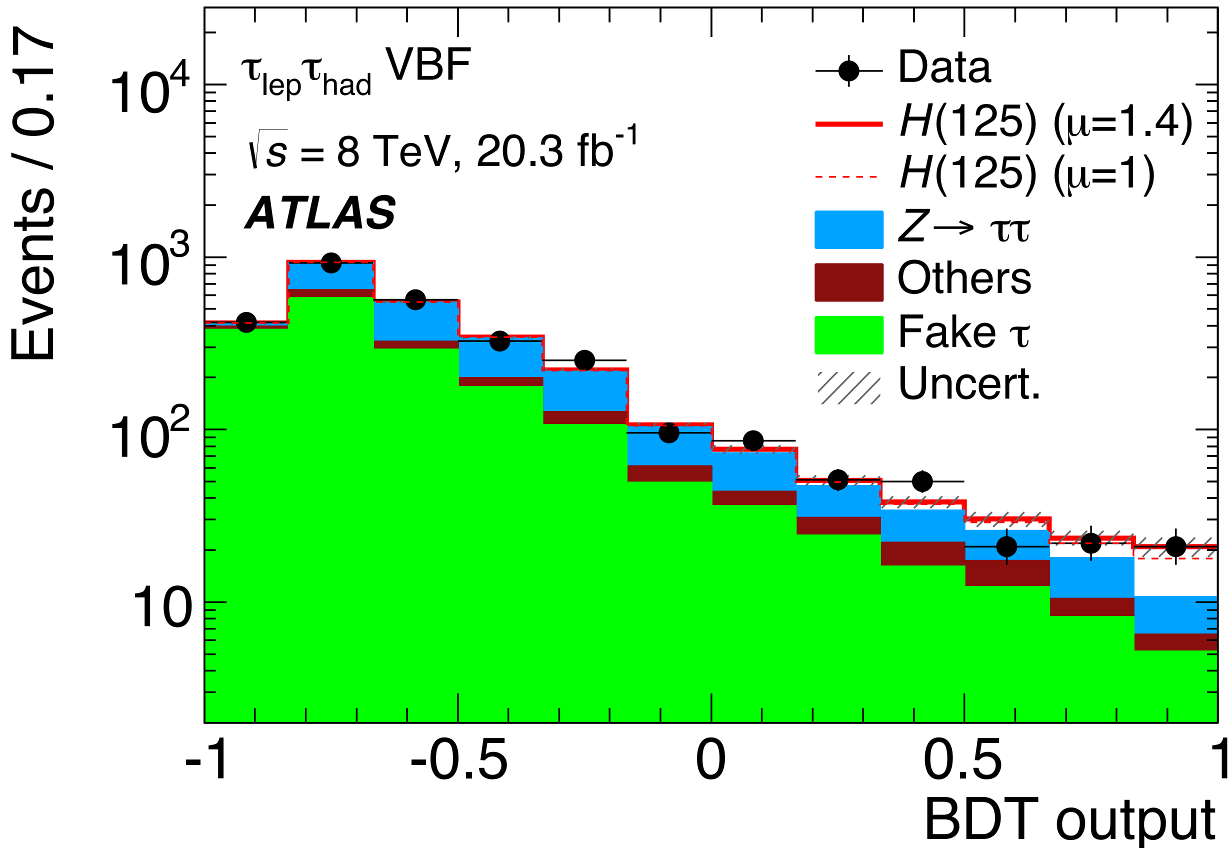


Case of multiple backgrounds

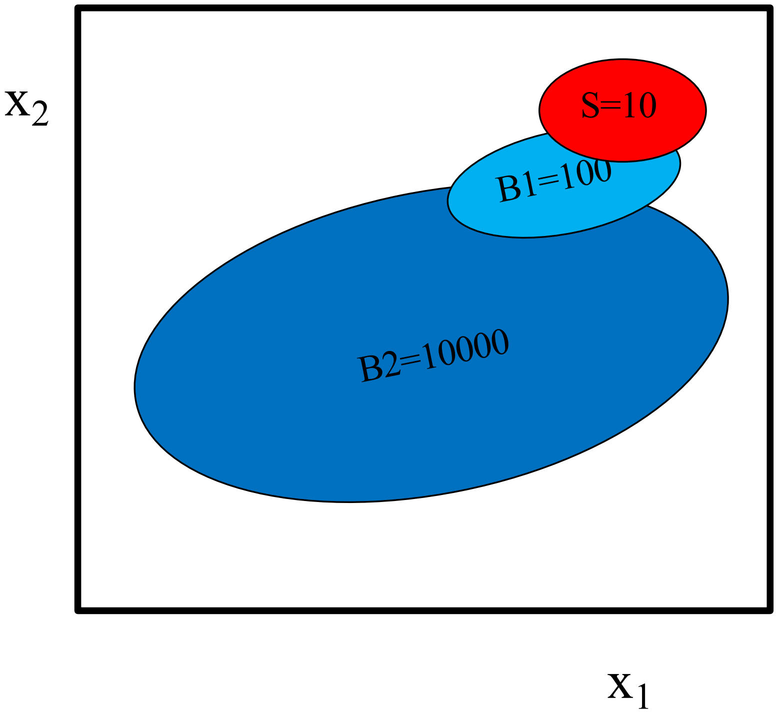
- Now suppose we have two different backgrounds:
- $b=b_1+b_2=L * \sigma_{bkg1}*\varepsilon_{bkg1} + L * \sigma_{bkg2}*\varepsilon_{bkg2}$
- $b=b_1+b_2=L * \sigma_{bkg1}* N_{pass1}/N_{total1} + L * \sigma_{bkg2}* N_{pass2}/N_{total2}$
- If I define the event weight
  - For dataset bkg 1 : $w_i= L * \sigma_{bkg1/} N_{total1}$
  - For dataset bkg 2 : $w_i= L * \sigma_{bkg2/} N_{total2}$
- And then : $b=\Sigma_{pass1}$ w_i + Σ_{pass2} w_i
- So I can merge both datasets and …
- $b= \Sigma_{pass\ 1\ and\ 2}\ w_i$
- ditto for many backgrounds… (effective for collaborative work)

Multiple backgrounds

- Such plots can be made directly



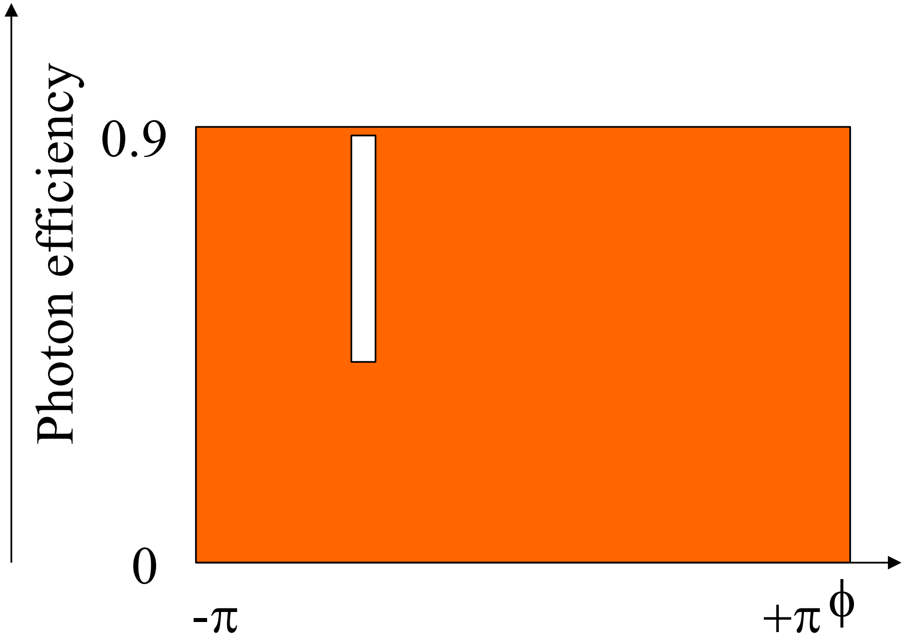
ML Application



- B1 is the more annoying background : smaller but more similar to Signal
- One can increase B1 dataset size and not B2, use weights for proper relative normalisation

Efficiency correction

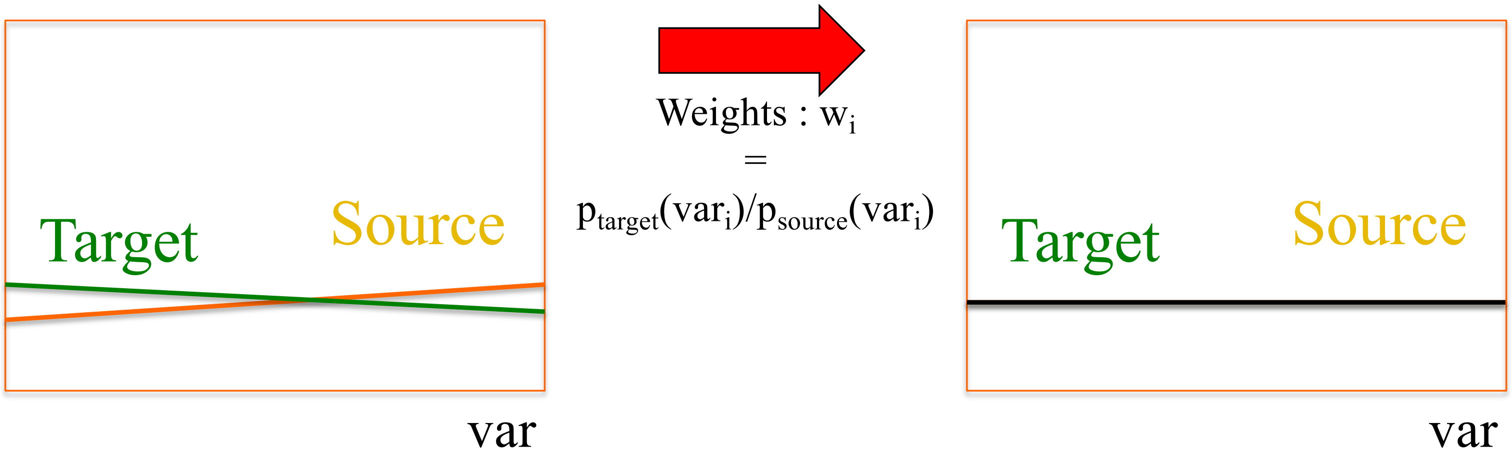
- Suppose, a detector malfunction causes photon efficiency to be halved in a small region of the detector
- ➔resimulate everything taking into account this effect ? (== billion of compute hours)



- No, simply define a new weight:
  - w_i^{photon}=0.5 if one photon in that region, 1 elsewhere
- Then w_i*= w_i^{photon} 🡐 weights of different sources can be multiplied
- And voilà, all event counting, all distributions are automagically corrected
- Particularly handy in large collaborations where many teams work on different aspect of event detection.
  - Each team comes up with its own weight
  - Physicists doing analysis can use (almost) blindly the weights they are given

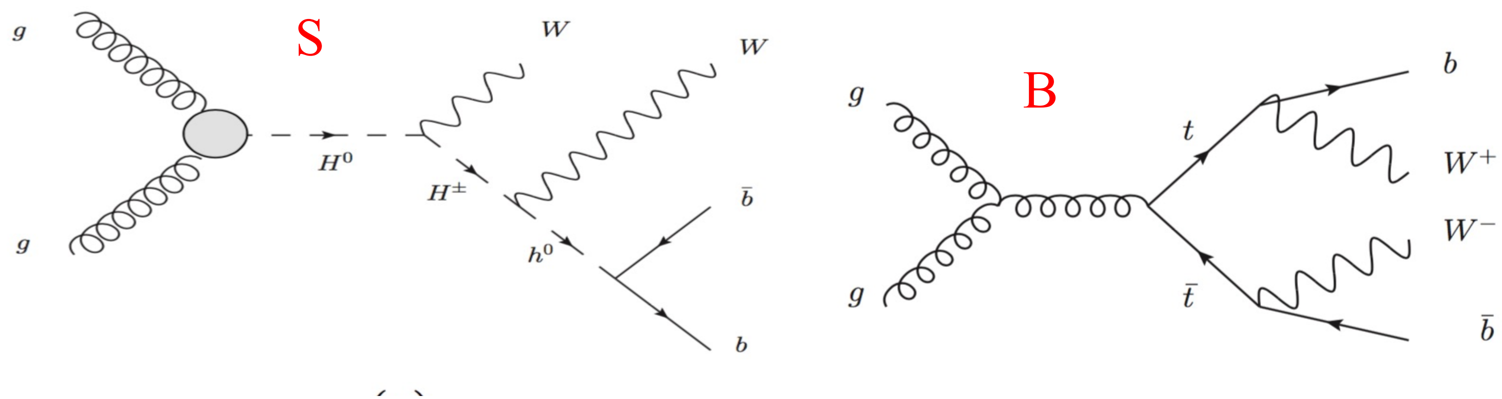
General Re-weighting

- Suppose a feature distribution is slightly different between a Source (e.g. Monte Carlo) and a Target (e.g. real data)
  - ➔reweight! …then use reweighted events



Event Generator weight

- Generators are software which creates event with multi particle final states with very precise correlation
- Very complex calculations
- ➔weighted events (weights can even be negative!)



Uncertainty

- When counting unweighted events uncertainty (Poisson case):
  - $N_{pass}=\Sigma_{pass}\ 1$
  - $\sigma\ N_{pass}=\sqrt{N_{pass}}=\sqrt{\Sigma_{pass}\ 1}$
  - $\sigma\ N_{pass}\ /\ N_{pass}=1/\sqrt{N_{pass}}$
- For weighted events (Poisson, binomial more involved):
  - $N_{pass}=\Sigma_{pass}\ w_i$
  - $\sigma\ N_{pass}=\sqrt{\Sigma_{pass}\ w^2_i}$
  - $\sigma\ N_{pass}\ /\ N_{pass}=\sqrt{\Sigma_{pass}\ w^2_i}\ /\Sigma_{pass}\ w_i$
  - Note : if $w_i=1$ ➔ like unweighted
  - Note : if I scale all weights by a : $\sigma\ N_{pass}\ /\ N_{pass}$ is unchanged (as expected)

power 2!!!

Effective number of events

- Suppose I have 2, and I add 1 (50%) in quadrature? What is the percentage increase ? (5 seconds)
- 12% ! $\sqrt{(2^2+1^2)}/2= \sqrt{5}/2=1.118$
- Meaning : quadratic sum is dominated by the largest values
- ➔having large weights destroy the statistical sensitivity
- Effective number of events of a sample == number of events of an equivalent weightless sample bringing the same precision
  - $N_{eff}= \Sigma^2 w_i / \Sigma w^2_i$
  - $N_{eff}/N=1/(1+Var(x)/<x>^2) <1$
  - The larger the distribution of weights the larger the loss of sensitivity

Caveats

- Reweighting applicable for small-ish corrections (otherwise variance of weight too large → loss of sensitivity)
- Of course cannot "invent" events
- Not really suitable to rescale variables (if says Energy of a particle is wrong by 2%, better rescale energy directly)
- Also weights are ~easy to compute if uncorrelated
- If correlated, can do 2-dimension reweighting more difficult (curse of dimensionality)
- Beware : not all software tools handle weights correctly, most tools do not handle negative weights correctly

Weights : w_i

$$= p_{target}(var_i)/p_{source}(var_i)$$