Signal / background preparation for (p)DNN Model Training

- We want to **prepare** the training set as in the HH->bbττ
- To **prepare** means:
  - 1) *Equalize statistics* of the total Signal sample (all masses) and total Background sample
    - NOTE 1 :
      - Beforehand we want to equalize signal samples corresponding to different mass values
    - NOTE 2:
      - In the total background sample, different processes contribute according to the relative importance according to their SM cross section
  - 2) *Apply variable scaling* considering Mean and RMS of the distribution obtained with each event contributing “as in the training set”
    - Therefore Mean and RMS must be computed on distributions built with event weights equal to those used for the training

Signal / background preparation for (p)DNN Model Training

Equalize statistics of total Signal and Background

Easy CASE: no intrinsic (physics and simulation related) weights;

1 single background B

- We have n signal masses, hence n signal samples, Sj, with j=1, .., n
- The number of events passing the preselection (AODmaker+any preselection applied on the trees) for Sj is Nj
- The number of background events = Nb

- Logic to define scale factors for the events (common to the entire sample)
  - Order according to increasing Nj the n signal samples
    - => $N_1 < N_2 < \ldots < N_n$
  - Compute the signal sample scale factors Fj = 1, N_1/N_2, N_2/N_3, … N_1/N_n
    - The stat.s of the entire signal sample is ~nN_1
  - If nN1 < Nb
    - apply to the background the scale factor Fb = nN_1/Nb
  - Else
    - Apply to signal events the scale factors Fj = Fj * Nb/(nN_1)

SIGNAL / BACKGROUND PREPARATION FOR (P)DNN MODEL TRAINING

Equalize statistics of total Signal and Background

Real CASE: intrinsic (physics and simulation related) weights are always considered, for signal and for background events

=> the total number of events* in the signal sample Sj is Nj = $\sum w_i$

=> the total number of background events* is Nb = $\sum_{process} (\sum w_i)$

- (*) = passing the preselection

- Logic to define scale factors for the events (common to the entire sample)
  - Order according to increasing Nj the n signal samples
    - => $N_1 < N_2 < \ldots < N_n$
  - Compute the signal sample scale factors Fj = 1, N_1/N_2, N_2/N_3, … N_1/N_n
    - The stat.s of the entire signal sample is ~nN_1
  - If nN1 < Nb
    - apply to the background the scale factor Fb = nN_1/Nb
  - Else
    - Apply to signal events the scale factors Fj = Fj * Nb/(nN_1)

As before

Signal / background preparation for (p)DNN Model Training

Real CASE

Equalize statistics of total Signal and Background

1) Whenever and event is used it comes with its intrinsic weight w_i

2) In the training and in the scaling any event is used with a training weight

$w_{Ti} = w_i F_{sample}$

Apply variable scaling

Produce distribution of each input variable using the training weights

=> compute Mean and RMS

Redefine (scale) the valiables of the training set