Spin & Systematics Status

J. Pela
Imperial College London

Spin Analysis Status

- New official samples processed into ntuples and now available.
  - Analysis code needed alterations to deal with several signal samples. (90% finished)
  - Several plots made to compare samples to be made available soon.
- Noticed a problems in the categorization boundaries.
  - Distributions sometime now cross more than once. (more than 2 bins likely needed)
  - Several alternative signal hypotesis (SM-like and alternative models/samples), so binning optimization should be redone for each “SM/alternative model” pair.
- **Automatic binning method should be implemented**

Proposed binning method

- To keep the analyze simple and based on the main Higgs Analysis, BDT boundaries will be the same as the ones of the main analysis
- Two or more categories will defined by maximizing:

$$\sum_{bins} \frac{\left|\left(entries_a - entries_b\right)\right|}{error_a + error_b}$$

- Process of rebining should be run again each new data milestone for each process sample pair

Proposal for Function Selection Systematics evaluations

- **Idea:** Use the data itself (binned or not) to generate toys and fit them.
- **A priori:**
  - Choose the categories of functions that would be acceptable to fit the background in the data.
  - Select degree of function to be used based on the predefined p-Value minimum p-Value of fit to data
- **Toy generation:**
  - This can be done in an un-binned way, directly from data, by integrating the mass spectrum to obtain the Cumulative distribution function of the data.
    - Will be a step function of nEvents-1 steps
  - This distribution can be made continuous in several ways (mid-step linearization, etc)

Proposal for Function Selection Systematics evaluations I

- **Idea:** Use the data itself (binned or not) to generate toys and fit them.
- **A priori:**
  - Choose the categories of functions that would be acceptable to fit the background in the data.
  - Select degree of function to be used based on the predefined p-Value minimum p-Value of fit to data
- **Toy generation:**
  - This can be done in an un-binned way, directly from data, by integrating the mass spectrum to obtain the Cumulative distribution function of the data.
    - Will be a step function of nEvents-1 steps
  - This distribution can be made continuous in several ways (mid-step linearization, etc)
  - The resulting distribution can be inverted and used to throw toys.

Proposal for Function Selection Systematics evaluations II

- Fit toys: with each one of the candidate functions to the background only side bands for each toy
- Plot resulting: Signal area background event prediction.
  - For the function pre-selected to fit the background on data (lowest p-Value on data) the sigma will be the statistical error of the fit.
  - The systematic error would be the highest difference between the fit to data prediction and central value of each of the test functions results over the toys.

Motivations

- Less complicated than the previous methods.
- This method will be sensitive to the movement of each function prediction due to low statistics oscillations, we will gain knowledge about how the predictions value will change and what is actually the most probable central for the know statistics.
- From the assumption that we would only use good/simple enough functions to fit (as opposite to use all possible functions) so a finite number of functions needs to be tested.
- Would be easy to repeat with each new dataset of data.

Questions

- This will obviously influenced by the statistical fluctuations already on data, but they will be smeared by the randomness of the method itself, is this desirable or a negative effect?
- This could be used to extract the statistical error of each function fit directly since no binning is implied. Does this provide extra information?